



La dimension argumentative des collocations textuelles en corpus électronique spécialisé au domaine du TAL(N)

Estelle Dubreil

► To cite this version:

Estelle Dubreil. La dimension argumentative des collocations textuelles en corpus électronique spécialisé au domaine du TAL(N). Linguistique. Université de Nantes, 2006. Français. NNT : . tel-00486063

HAL Id: tel-00486063

<https://theses.hal.science/tel-00486063>

Submitted on 24 May 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Université de Nantes

ÉCOLE DOCTORALE CLC

« CONNAISSANCE LANGAGES CULTURES »

Année 2006

N° attribué par la bibliothèque

--	--	--	--	--	--	--	--	--	--

La dimension argumentative des collocations textuelles en corpus électronique spécialisé au domaine du TAL(N)

THÈSE

pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ DE NANTES

Discipline : SCIENCES DU LANGAGE

présentée et soutenue publiquement par

Estelle DUBREIL

le 20 octobre 2006

à l'UFR Sciences et Techniques, Université de Nantes

devant le jury ci-dessous

Président :	Geoffrey WILLIAMS, Professeur des Universités	ADICORE, Univ.Bretagne
Rapporteurs :	Alain POLGUÈRE, Professeur des Universités	OLST, Univ.Montréal
	François RASTIER, Directeur de recherche - CNRS	MoDyCo, Univ.Paris X
Examineurs :	Béatrice DAILLE, Professeur des Universités	LINA, Univ.Nantes
	Olga GALATANU, Professeur des Universités	CERCI, Univ.Nantes
	Marie-Paule PÉRY-WOODLEY, Professeur des Universités	ERSS, Univ.Toulouse 2

Directeurs de thèse : Professeur Béatrice DAILLE – Professeur Olga GALATANU

Laboratoire : Laboratoire d'Informatique de Nantes Atlantique (LINA) CNRS-FRE 2729

LA DIMENSION ARGUMENTATIVE DES
COLLOCATIONS TEXTUELLES EN CORPUS
ÉLECTRONIQUE SPECIALISÉ AU DOMAINE DU
TAL(N)

*The Argumentative Part of Textual Collocations in an
Electronic Corpus specialized on the TAL(N) Community*

Estelle DUBREIL



favet neptunus eunti

Université de Nantes

Estelle DUBREIL

*La dimension argumentative des collocations textuelles en corpus électronique
spécialisé au domaine du TAL(N)*

xii+302.

– *Plus on s'approche de son rêve, plus la Légende Personnelle devient la véritable raison de vivre.*

L'Alchimiste, p. 92 — Paulo COELHO.

– *Une quête commence toujours par la Chance du Débutant. Et s'achève toujours par l'Epreuve du Conquérant.*

L'Alchimiste, p. 155 — Paulo COELHO.

Remerciements

Je remercie Béatrice DAILLE. Béa, merci d'abord d'avoir accepté la codirection de cette thèse et par là d'avoir tenté l'aventure fructueuse d'une cohabitation informaticiens-linguiste. Merci ensuite de m'avoir si bien intégrée à l'équipe LN par mes présentations aux séminaires, de m'avoir chargée de la communication à l'occasion de TALN 2003, de m'avoir associée au projet DECO, de m'avoir proposé un travail de coencadrement... Toutes ces expériences fondent ma volonté farouche d'exercer un jour le métier d'enseignant-chercheur. Merci enfin pour la qualité de ton encadrement fondamentalement humain, basé sur la disponibilité et la confiance, fait d'une succession équilibrée de rigueur et de souplesse.

Je remercie Olga GALATANU. Olga, vous êtes d'abord à l'origine de mon intérêt pour la sémantique, la pragmatique et l'analyse du discours. Vous avez ensuite joué un rôle important dans le financement de cette thèse, si nécessaire à une disponibilité intellectuelle et professionnelle et à la stabilité d'un équilibre personnel. Enfin, le positivisme qui vous caractérise se révèle une qualité salubre à si longue échéance ; merci d'avoir si souvent trouvé ces problématiques de recherche « forrrmidables » !

Je remercie Geoffrey WILLIAMS, Alain POLGUÈRE, François RASTIER et Marie-Paule PÉRY-WOODLEY pour avoir accepté de participer à ce jury de thèse.

Je remercie Geoffrey WILLIAMS. « You shall know a [friend] by the company it keeps! » Merci de l'intérêt que tu as porté à cette recherche dès son commencement. Merci également d'en avoir facilité l'issue par de nombreux échanges enrichissants, par les articles mis à ma disposition, par les éclaircissements apportés sur les pratiques de corpus, par la participation aux journées de linguistique de corpus et à Euralex 2004.

Je remercie Jean VÉRONIS pour m'avoir permis à plusieurs reprises d'améliorer la performance de mes évaluations en me donnant *les clés* du problème, évitant ainsi bien des erreurs dans les résultats. Question de « Contextes » !

Je remercie Frédéric BENHAMOU pour son accueil enthousiaste au sein du LINA. Outre le besoin en contacts humains nécessaires à la réussite d'une telle entreprise, ce rattachement au laboratoire d'informatique m'a apporté un soutien matériel qui, loin d'être trivial, fonde la possibilité même de prétendre à l'efficacité.

Je remercie Nordine FOUROUR pour son implication professionnelle et personnelle fondamentale aux balbutiements de cette thèse. Nordine, je te retourne ce compliment : merci pour ton amitié précieuse.

Je remercie Jim NORMAND, sans qui cette thèse aurait probablement eu du mal à se terminer dans les temps impartis, eu égard à quelques problèmes annexes, abstract-ion faite de la bibliographie ! Merci pour ta disponibilité et tes encouragements quotidiens, en particulier les dernières semaines.

Je remercie Philippe GÉRARD, qui sans avoir jugé du neuf a corrigé bien des coquilles... Philippe, merci pour ton professionnalisme dans le travail de correction du tapuscrit et pour ta capacité à anticiper sur des échéances pourtant courtes !

Je remercie Fabrice EVEN et Alexandra REYNOLDS, en résumé pour m'avoir fait si bonne impression !!!

Sans changer de registre, je remercie Lorraine GOEURIOT et Emmanuel MORIN, de l'équipe LN, dont la compagnie en contexte professionnel est réellement enrichissante.

Je remercie Virginie BRAUD, dont l'amitié sincère est propice à l'écoute et à l'échange, dans toute la circularité qu'autorise l'interaction verbale.

Je remercie Brice PAJOT pour sa présence rassurante, souvent même réconfortante, dans le quotidien de l'exercice difficile que représente la phase de rédaction de thèse.

Je remercie tous « mes collègues et néanmoins amis » du LINA : Jérémie BOURDON, Jean BÉZIVIN, Marc CHRISTIE et Nathalie DENIS, Guillaume FERTIN et Hélène FARINE, Gaëtan GAUMER et Anne, Frédéric GOUALARD et Magali HERSANT, Freddy PERRAUD, Guillaume RASCHIA et Murielle PROVOST, Dalila TAMZALIT, Charlotte TRUCHET... envers qui ma seule force a bien souvent été de confier mes moments de faiblesse. Merci à tous pour ces échanges informels et ô combien réconfortants, glanés aux détours des couloirs.

Je remercie « toute *ma* famille » : Maman, Papa, Mimy et Marc, Ludy et Nico, pour leur soutien affectif, moral et surtout respectueux tout au long de ce parcours.

Je remercie également Frédéric DEGOUZON et Hélène CUVILLIER, pour leur curiosité et leurs encouragements hebdomadaires dans l'avancement de ce travail.

Mon ALEX, tu es le dernier que je remercie, tu es le plus petit aussi, et pourtant...
Par la fierté que tu m'as tant de fois témoignée, par cette thèse à demi commencée,
Sur « les Trésors des Pirates et du monde entier »
Tu as su bien souvent piquer mon orgueil,
Sentiment nécessaire à l'aboutissement du travail sans écueil,
Du haut de ta petite enfance, j'ai redécouvert la persévérance. Merci, mon chéri.

Introduction	1
Chapitre 1 – Collocations : définitions et problématiques	7
Présentation.....	7
1 Les phénomènes phraséologiques.....	8
1.1 Définition de la phraséologie	6
1.2 Présentation des phénomènes.....	8
2 Les phénomènes collocationnels	10
2.1 Problèmes de définition	10
2.1.1 Une nature dynamique	10
2.1.2 Un continuum libre – figé.....	10
2.1.3 Une approche prototypique	11
2.2 Solutions de description	12
2.2.1 Critères de figement	12
2.2.1.1 Critère de la transparence.....	12
2.2.1.2 Critère de l’arbitraire.....	12
2.2.2 Critères définitoires	13
2.2.2.1 Critères de Tutin et Grossmann	13
2.2.2.2 Critères de Williams	15
2.3 Deux approches du concept de « collocation »	16
2.3.1 Approche lexicologique – lexicographique	16
2.3.1.1 Définition restreinte	16
2.3.1.2 Typologies.....	16
2.3.2 Approche « linguistique de corpus »	19
2.3.2.1 Historique	19
2.3.2.2 Définition large	20
2.3.2.3 Typologies.....	21
3 Les problématiques en linguistique appliquée	24
3.1 La didactique des langues	24
3.2 La lexicographie	25
3.3 La terminologie	27
4 Les problématiques en linguistique théorique.....	27
4.1 La modélisation	27
4.1.1 Pourquoi et comment modéliser les collocations ?.....	27
4.1.2 Le modèle des fonctions lexicales – FL.....	28
4.1.2.1 Fonctions lexicales paradigmatiques <i>vs</i> syntagmatiques	29
4.1.2.2 Fonctions lexicales standard <i>vs</i> non standard.....	29
4.1.2.3 Fonctions lexicales simples <i>vs</i> complexes <i>vs</i> configurations de FL.....	30 34

4.1.2.4	Limites du modèle des fonctions lexicales	30
4.2	La motivation sémantique	31
4.2.1	Les procédés sémantiques productifs.....	31
4.2.2	Les régularités sémantiques.....	33
5	Une motivation argumentative.....	35
5.1	Problématique et filiation de la recherche	35
5.2	Définition retenue	35
5.2.1	Les collocations textuelles.....	35
5.2.2	Les collocations libres <i>vs</i> restreintes.....	36
	Conclusion.....	37

Chapitre 2 – La sémantique des possibles argumentatifs 39

	Présentation	39
1	La sémantique argumentative	40
1.1	La théorie de la polyphonie	40
1.2	L'argumentation dans la langue	40
1.3	La théorie des topoï.....	43
1.3.1	Principes topiques de l'argumentation	43
1.3.2	Propriétés des topoï.....	44
1.3.3	Topoï intrinsèques et extrinsèques.....	44
1.4	Les blocs d'argumentations.....	46
1.4.1	Préceptes de la théorie des blocs d'argumentations	46
1.4.2	Argumentation externe.....	47
1.4.3	Argumentation interne.....	47
2	La sémantique des possibles argumentatifs	48
2.1	Filiations et postulats	48
2.1.1	Filiations logico-structuraliste et historico-psychologique.....	48
2.1.2	Aspect tridimensionnel de la signification.....	49
2.1.2.1	Une approche dénotative et référentielle.....	49
2.1.2.2	Une approche argumentative et encyclopédique	50
2.1.2.3	Une approche associative et holistique.....	51
2.2	Représentation de la signification lexicale	52
2.2.1	Les conditions nécessaires	52
2.2.2	Le noyau et les stéréotypes	52
2.2.3	Les possibles argumentatifs.....	53
2.2.3.1	Les valeurs axiologiques	54
2.2.3.2	Les mécanismes de l'axiologisation.....	55
2.2.4	Le « monde quantique » de la signification	56
2.3	Applications et interactions.....	57
2.3.1	Considérations méthodologiques	57
2.3.2	Analyse linguistique du discours	59
	Conclusion.....	60

Chapitre 3 – Principes de constitution du corpus TAL(N)	61
Présentation.....	61
1 « corpus »	62
1.1 Le concept général.....	62
1.2 Le corpus électronique.....	64
1.2.1 Le corpus informatisé	64
1.2.2 Le corpus de référence	65
1.2.3 Le corpus spécialisé	66
2 Circonscrire un domaine	68
2.1 Langue spécialisée vs langue de spécialité	68
2.2 Théorie des sous-langages.....	69
2.3 Théorie des communautés de discours	71
2.4 Justification du choix théorique	73
3 Constitution du corpus TAL(N)	74
3.1 Délimitation du domaine scientifique TAL	74
3.1.1 Le domaine pressenti.....	74
3.1.2 La communauté de discours TAL	74
3.1.2.1 Un but commun.....	75
3.1.2.2 Des mécanismes d'interaction.....	75
3.1.2.3 Des mécanismes de participation.....	75
3.1.2.4 L'utilisation et la possession d'un ou de plusieurs genres	75
3.1.2.5 L'acquisition d'un lexique spécifique.....	76
3.1.2.6 Un seuil de membres.....	77
3.1.3 Une communauté de domaine	77
3.2 Sélection des données textuelles	78
4 Construction du corpus TAL(N).....	79
4.1 Méthodologie d'obtention du corpus	79
4.1.1 Les fichiers.....	79
4.1.2 Le nettoyage.....	80
4.1.3 Les traitements.....	81
4.2 Résultat.....	82
Conclusion.....	82

Chapitre 4 – Hypothèses de recherche et protocole de validation	83
Présentation.....	83
1 Hypothèses de recherche	84
1.1 Première hypothèse – H1	84
1.2 Deuxième hypothèse – H2	84
1.3 Troisième hypothèse – H3	85
1.3.1 Collocation argumentative – H31	85
1.3.2 Motivation argumentative – H32.....	88
1.3.3 Collocation arbitraire – H33	90
1.3.4 Quatrième hypothèse – H4	91
2 Protocole de validation	93
2.1 Linguistique de corpus	93
2.1.1 Le repérage des unités lexicales	93

2.1.1.1	Démarche déductive	93
2.1.1.2	Démarche inductive.....	96
2.1.2	Dépouillement des données	98
2.1.2.1	Le critère statistique	98
2.1.2.2	Le critère de pertinence.....	98
2.1.2.3	Le critère lexical.....	99
2.1.2.4	Le critère terminologique	99
2.2	Lexicographie.....	100
2.2.1	Les dictionnaires Larousse Lexis et Grand Robert.....	100
2.2.1.1	Objectif commun	100
2.2.1.2	Traditions différentes	101
2.2.1.3	Nomenclature lexicographique.....	103
2.2.2	Principes de construction de la signification lexicale	105
2.2.2.1	Les traits nécessaires de catégorisation	105
2.2.2.2	Le noyau	105
2.2.2.3	Les stéréotypes	108
2.2.2.4	Les possibles argumentatifs.....	108
2.3	Validation des données	109
2.3.1	Le test de substitution argumentatif.....	109
2.3.2	Le test d'enchaînement argumentatif.....	110
2.3.3	Considérations méthodologiques	111
2.3.3.1	Les tableaux de données.....	111
2.3.3.2	Synthèse	112
Conclusion	112

Chapitre 5 – Analyse sémantico-discursive des collocations textuelles 113

Présentation	115
1 Les monovalents positifs.....		116
1.1 Améliorer – Amélioration(s)		116
1.1.1 Dépouillement des données		116
1.1.1.1 Fréquences et nettoyage.....		116
1.1.1.2 Liste des collocations lexicales.....		116
1.1.2 Significations lexicales de améliorer – amélioration		118
1.1.2.1 Signification lexicale de améliorer		118
1.1.2.2 Signification lexicale de amélioration		119
1.1.3 Analyse des données		120
1.1.3.1 Le collocatif est un nom		120
1.1.3.2 Le collocatif est un adjectif		125
1.2 Performance(s)		128
1.2.1 Dépouillement des données		128
1.2.1.1 Fréquences et nettoyage.....		128
1.2.1.2 Liste des collocations lexicales.....		128
1.2.2 Signification lexicale de performance		130
1.2.3 Analyse des données		132
1.2.3.1 Le collocatif est un nom		132
1.2.3.2 Le collocatif est un adjectif		139
1.3 Synthèse sur les monovalents positifs		144

2	Les monovalents négatifs	147
2.1	Erreur(s).....	147
2.1.1	Dépouillement des données.....	147
2.1.1.1	Fréquences et nettoyage	147
2.1.1.2	Liste des collocations lexicales	148
2.1.2	Signification lexicale de erreur.....	149
2.1.3	Analyse des données.....	150
2.1.3.1	Le collocatif est un nom.....	150
2.1.3.2	Le collocatif est un adjectif.....	156
2.2	Problème(s)	158
2.2.1	Dépouillement des données.....	158
2.2.1.1	Fréquences et nettoyage	158
2.2.1.2	Liste des collocations lexicales	158
2.2.2	Signification lexicale de problème.....	160
2.2.3	Analyse des données	161
2.2.3.1	Le collocatif est un nom.....	161
2.2.3.2	Le collocatif est un adjectif.....	166
2.3	Synthèse sur les monovalents négatifs.....	173
3	Les bivalents	176
3.1	Évaluer – Évaluation(s)	176
3.1.1	Dépouillement des données.....	176
3.1.1.1	Fréquences et nettoyage	176
3.1.1.2	Liste des collocations lexicales	177
3.1.2	Signification lexicale de évaluer – évaluation	179
3.1.2.1	Signification lexicale de évaluer	179
3.1.2.2	Signification lexicale de évaluation	180
3.1.3	Analyse des données.....	182
3.1.3.1	Le collocatif est un nom.....	182
3.1.3.2	Le collocatif est un adjectif.....	191
3.2	Résultat(s).....	198
3.2.1	Dépouillement des données.....	198
3.2.1.1	Nettoyage.....	198
3.2.1.2	Liste des collocations lexicales	198
3.2.2	Signification lexicale de résultat	201
3.2.3	Analyse des données.....	202
3.2.3.1	Le collocatif est un nom.....	202
3.2.3.2	Le collocatif est un adjectif.....	209
3.3	Synthèse sur les bivalents.....	216
4	Synthèse générale	219
	Conclusion.....	221

Conclusion et perspectives **223**

Bibliographie **231**

Annexes	247
Annexe 1 (Chap. 3) : données sur la communauté TAL.	248
Annexe 2 (Chap. 3) : liste de référencement des articles du corpus.	251
Annexe 3 (Chap. 4) : liste des concordances droites et gauches de performance-s.	261
Annexe 4 (Chap. 5) : articles lexicographiques de améliorer – amélioration.	288
Annexe 5 (Chap. 5) : articles lexicographiques de performance.	290
Annexe 6 (Chap. 5) : articles lexicographiques de erreur.	292
Annexe 7 (Chap. 5) : articles lexicographiques de problème.	296
Annexe 8 (Chap. 5) : articles lexicographiques de évaluer – évaluation.	298
Annexe 9 (Chap. 5) : articles lexicographiques de résultat.	300

Introduction

Préambule

« Corpus linguistics does not exist » (Chomsky, entretien avec Aarts [Rastier 2005, p. 40]). De façon radicale, une certaine linguistique fondamentale, en particulier générative, s'oppose à l'utilisation de corpus attestés, dont les données entravent l'accès au système de la langue. « En corpus, la grammaire universelle se trouve souillée par la culture, la société, l'humeur ou les pathologies du locuteur, les choix, la sélection de l'analyste, etc. » [Mayaffre 2005, p. 7]. Plus raisonnablement, longtemps le recours aux corpus fut fonction de l'objet d'étude du chercheur, selon que celui-ci s'intéresse à la langue ou à la parole, dualité forte héritée de Saussure. Naturellement, dans la perspective d'une étude phonologique, syntaxique ou sémantique centrée sur la compétence, l'intérêt porté aux données attestées était inexistant, relégué aux disciplines dont l'objet d'étude relève de la performance, à commencer par la pragmatique, l'analyse du discours mais aussi la stylistique et la littérature, par exemple. Or l'époque d'une linguistique sans corpus semble aujourd'hui révolue. Toutes les disciplines utilisent des corpus, on parle de corpus en sémantique, en phonologie, de corpus de traductologie, etc., et le terme *corpus* est amené à désigner des réalités différentes : on parle de corpus construit, corpus attesté, mais aussi de corpus de référence, corpus spécialisé, etc. La variété des emplois trahit la variété des domaines et des pratiques qui en sous-tendent l'utilisation. « Si tout le monde conçoit désormais que le corpus est un *observable* nécessaire en linguistique, au moins deux approches se font face pour peut-être se compléter. Pour les uns, le corpus est un *observatoire* d'une théorie *a priori*, pour les autres, le corpus est un *observé dynamique* qui permet de décrire puis d'élaborer des modèles *a posteriori*. Théorie et empirie, déduction et induction, linguistique de la langue et linguistique de la parole..., en ce moment, l'épistémologie fondamentale de la discipline se joue et se rejoue, parfois avec naïveté, parfois avec force, dans la réflexion sur les corpus. » [Mayaffre 2005, p. 6]

On distingue habituellement trois catégories de corpus. Les corpus lexicographiques ou *sacs de mots*, généralement fermés, ont la particularité de pouvoir prétendre à l'exhaustivité. Les corpus phrastiques, dédiés aux pratiques syntaxiques ou sémantiques, sont construits pour éprouver la validité d'une hypothèse théorique. Enfin, les corpus textuels, composés d'*archives de textes*, sont nécessairement constitués de données attestées, « puisqu'on ne saurait fabriquer artificiellement un texte pour prétendre en appréhender le sens » [Mayaffre 2005, p. 11]. Le corpus textuel est aujourd'hui de nature électronique. Aussi, le développement d'outils informatiques adaptés à l'exploration de ces données textuelles et les analyses qui en résultent ont été à la base d'une nouvelle discipline : la linguistique de corpus. Cette discipline, depuis ses prémices dans le monde anglo-saxon en 1970 [Sinclair et al. 1970, 2002], mène une réflexion sur la pratique de corpus, en cherchant à faire de celui-ci un objet d'étude à part entière. Entre linguistique et informatique, la linguistique de corpus pose la question du rapport entretenu avec la recherche d'attestations. D'un point de vue technique, traiter une grande quantité de données offre de nouvelles possibilités d'analyses. D'ailleurs, on observe depuis quelques années une fusion progressive de la linguistique de corpus avec le domaine du TAL [Daille et Romary 2001]. Toutefois, selon Rastier, pour la

linguistique de corpus « l’informatique reste un instrument, non un modèle théorique, car la linguistique appartient pleinement aux sciences de la culture » [Rastier 2005, p. 41]. En outre, contrairement aux idées reçues, la linguistique de corpus ne pratique pas d’analyse automatique des données, au sens où le choix raisonné de textes qui préexiste à la conception de tout corpus et l’interprétation des données qui succède à toute analyse impliquent l’intervention du chercheur : « Les données sont ce qu’on se donne. » [Rastier 2005]. D’un point de vue empirique, les observables produits par la linguistique de corpus sont nouveaux, au sens où les faits observés ont l’apanage de l’objectivité et revêtent à ce titre une portée scientifique. En ce sens, ils reposent la question du rapport entre empirie et théorie, tant en amont du processus de recherche dans la formulation des hypothèses qu’en aval dans la recherche de régularités et/ou de singularités. Cette approche empirique a permis de remplacer l’introspection fondée sur l’intuition du chercheur par une approche inductive [Tognini-Bonelli 2001, p. 3]. En outre, la linguistique de corpus fonde sa démarche sur l’analyse des textes, objet d’étude privilégié du linguiste : « L’objet empirique de la linguistique de corpus est constitué de textes (oraux ou écrits), non de mots ou de phrases, qui ne s’observent pas à l’état isolé et qui même lorsqu’ils le sont restent toujours relatifs à un genre et à un discours. » [Rastier 2005, p. 31]. Or l’émergence de régularités textuelles est propre à constituer – à recréer une norme linguistique, sociale, culturelle... Aussi, l’observation des données textuelles attestées abolit la dichotomie trop stricte entre langue et parole : « Il y a en effet actuellement une affirmation existentialiste ou phénoménaliste, qui consiste à refuser la dualité saussurienne comme dichotomie pour la concevoir comme une dyade, car il ne saurait y avoir d’essence sans existence, de système sans actualisation. » [Mayaffre 2005, p. 9]. Ce qui constitue d’un point de vue épistémologique une nouvelle approche.

Prises séparément, ni la sémantique ni l’analyse de corpus ne constituent d’approches linguistiques réellement nouvelles ; l’intérêt des problématiques les plus récentes repose précisément sur la combinaison de ces deux sphères, propre à confronter n’importe quel modèle théorique visant la mise à jour du « système » avec la réalité de la variation langagière. « En comparaison avec une linguistique uniquement introspective, la prise en compte des corpus en sémantique suppose en effet une rencontre avec un principe de réalité qui, tenant compte de la difficulté de constituer un corpus, de la résistance à la systémisation que présentent parfois les faits langagiers et du travail long que suppose l’élaboration d’une interprétation, s’oppose souvent à la vision d’un locuteur “ idéal ” qui permettrait de décrire un modèle stable, contrôlé et prédictif. » [Condamines 2005, p. 16]. Pour opposées que soient ces deux approches, leur combinaison, reposant sur l’analyse de données attestées, réinvente le *rapport* entre signification et sens par l’intermédiaire du discours et de sa norme, rendus accessibles par le corpus textuel. De fait, « les mots ont peut-être une signification en langue mais ils n’ont de sens et de valeur qu’en contexte » [Mayaffre 2005, p. 8].

Cadre théorique

Dans la perspective d’une combinaison sémantique – analyse de corpus, le modèle lexical de la sémantique des possibles argumentatifs – SPA – présente un intérêt certain, dans la mesure où la signification lexicale y est décrite de manière pluridimensionnelle, au sens où cette description intègre simultanément des informations sémantiques internes à la langue et des informations discursives externes donc contextuelles [Galatanu 1999a, 1999b, 2005a]. À

l'interface entre la sémantique logique et structuraliste d'un côté et la sémantique historico-psychologique et cognitive de l'autre, la SPA réconcilie ces deux traditions linguistiques en adhérant à un processus de stabilisation du monde dans et par la langue. Ce processus est simultanément subjectif et intersubjectif, au sens où il est inévitablement variable dans l'espace individuel et social [Galatanu 2000b], donc marqué d'un flou cinétique.

Concrètement, au niveau sémantique la description de la signification lexicale associe deux éléments : un noyau à caractère référentiel et des stéréotypes à caractère argumentatif, associés durablement aux mots. Au niveau discursif, la signification lexicale est orientée vers le potentiel discursif des mots, fondamentalement argumentatif, par le biais de possibles argumentatifs correspondant à l'association du mot avec l'un de ses stéréotypes dans le discours. Le modèle SPA assigne donc à la signification lexicale un aspect tridimensionnel, à la fois dénotatif, argumentatif et par répercussion associatif.

Cette approche SPA permet, au niveau sémantique, de mettre à jour des mécanismes explicatifs quant au cinétisme de la signification lexicale et par là de rendre compte du pouvoir qu'a le discours, de véhiculer des systèmes de valeur, de les transformer, mais aussi de les produire. Aussi, au niveau discursif la SPA fonctionne en interaction nécessaire avec l'analyse linguistique du discours – ALD –, dont les préoccupations portent sur les mécanismes de renforcement, d'activation, d'affaiblissement, d'interversion, etc. des valeurs portées par les unités lexicales dans le discours. Par voie de conséquence, l'ALD permet également de formuler des hypothèses interprétatives sur la fonction sociale du discours étudié et sur la reconstruction discursive des systèmes de valeurs. La SPA-ALD inscrivant son objet d'étude dans une praxéologie nécessaire, toute analyse menée repose sur des textes permettant l'accès à un discours représentatif d'une communauté, d'un domaine, etc.

Objet d'étude

L'analyse des phénomènes collocationnels est historiquement liée au développement de la linguistique de corpus. En réaction contre les courants linguistiques structuralistes, qui évacuent la question du sens des mots, l'approche collocationnelle prônée par Firth est fondamentalement pragmatique : elle place l'étude de la langue en contexte. Dans cette perspective, l'analyse des collocations participe de l'élaboration d'une théorie du sens contextualiste à l'origine de la linguistique de corpus. Aussi, lorsque Firth introduit pour la première fois le terme *collocation*, il entend caractériser certains phénomènes linguistiques de cooccurrence qui engendrent des modifications de sens constatées sur les mots par leur relation de collocation en contexte. Le contexte est accessible par le texte, donc, l'analyse des collocations se fonde sur l'observation des relations sémantiques créées par le contexte, à la surface des textes.

À l'interface entre discours et langue, par leur instanciation contextuelle et par leur caractère idiomatique, les collocations posent certaines difficultés de définitions. Les solutions de descriptions proposées s'apparentent à une série de critères discutés et discutables selon l'approche choisie. Outre les variations de définitions et la diversité des critères permettant de les décrire, les collocations sont issues des différentes méthodes de traitement et d'analyse des phénomènes, qui varient elles-mêmes selon les problématiques posées, les domaines et la finalité des études. Qu'il s'agisse de problématiques appliquées au domaine didactique, lexicographique, terminologique ou de problématiques théoriques, les objectifs convergent : il

s'agit d'une part de chercher à établir une modélisation de l'ensemble des phénomènes collocationnels, c'est-à-dire de mettre au point un modèle simultanément définitoire, descriptif et prédictif, et d'autre part de mettre à jour les mécanismes inhérents à la formation des collocations. De fait, la question de la motivation notamment sémantique des collocations est souvent évacuée dans les modèles lexicaux, tout en étant pourtant centrale en linguistique.

Problématique

Adhérant au point de vue de Tutin et Grossmann [Tutin et Grossmann 2002] concernant la nécessité d'étudier les mécanismes permettant de générer les collocations, l'idée de décrire argumentativement la motivation des entités mobilisées dans certains phénomènes collocationnels provient du pouvoir explicatif et prédictible « potentiel » de la sémantique des possibles argumentatifs, en lien avec la conception-description pluridimensionnelle de la signification lexicale.

Aussi, l'hypothèse particulière à cette recherche et plus fondamentalement encore la thèse défendue dans ce travail prennent la forme d'une intuition issue du rapprochement des définitions respectives de *collocations* et de *possibles argumentatifs*. Nous postulons en effet qu'une collocation correspond potentiellement à la concrétisation d'un possible argumentatif – PA – dans le discours. Cette intuition se décline sous la forme d'une triple hypothèse :

- soit la collocation repérée est la concrétisation d'un PA dans le discours ; dans ce cas, la *collocation* est qualifiée d'*argumentative* ;
- soit la collocation repérée est argumentativement motivée ; dans ce cas, le collocatif entretient une *relation argumentative* avec la base ;
- soit la collocation repérée ne correspond ni à la concrétisation d'un PA dans le discours ni à une association argumentativement motivée ; dans ce cas, la *collocation* est qualifiée d'*arbitraire*.

Nous avons créé à dessein le corpus électronique spécialisé TAL(N), représentatif de la communauté du traitement automatique du langage naturel, pour tester cette hypothèse. Le choix d'un corpus spécialisé se justifie à trois niveaux :

- d'abord au niveau SPA, pour éprouver le fondement de la théorie en lien avec de la subjectivité discursive inscrite dans les possibles argumentatifs des unités lexicales qui composent tout discours ; cette perspective implique d'ailleurs de rappeler préalablement l'objectivisme supposé du discours scientifique spécialisé, postulat dont nous avons pleinement conscience qu'il est stéréotypé ;
- ensuite au niveau de notre objet d'étude même, puisque dans la perspective contextualiste les études ont montré que les collocations assuraient un rôle dans la *cohésion textuelle* ; en ce sens, elles sont susceptibles de pouvoir rendre compte des régularités et/ou singularités du corpus textuel constitué et du discours et du registre auquel ce dernier réfère ;

- enfin au niveau ALD, pour mettre à jour les mécanismes discursifs propres au discours scientifique spécialisé TAL(N), dont nous pensons intuitivement qu'il « ne joue pas » à l'intervention des valeurs axiologiques, eu égard à un contrat de transparence et de non-ambiguïté dans la transmission du savoir. Nous nous attendons donc à trouver une majorité de potentiels activés, voire renforcés, en particulier pour les collocations arbitraires, mais pas ou peu de potentiels neutralisés ou affaiblis et encore moins de potentiels intervertis ou transgressés.

La démarche que nous avons présentée ci-dessus et que nous allons suivre tout au long du manuscrit est détaillée dans le plan du document proposé, par la présentation de chaque intitulé de chapitre et de son contenu.

Plan du document

Le chapitre 1, intitulé « *Collocations, définitions et problématiques* », est consacré aux phénomènes collocationnels. Il en présente et explique les problèmes de définitions et les solutions de descriptions fondées sur des critères et des typologies variables selon l'approche lexicologique-lexicographique ou linguistique de corpus retenue. Ces variations dans l'acceptabilité du concept sont également le fruit des différences de méthodes de traitement et d'analyse des phénomènes, qui varient elles-mêmes selon les problématiques appliquées ou théoriques posées, les domaines et la finalité des études. Dans cette optique, le chapitre se clôt sur l'étude de la motivation sémantique et sur notre proposition de description de l'association collocationnelle en termes de motivation argumentative.

Le chapitre 2, intitulé « *La sémantique des possibles argumentatifs* », a pour vocation de présenter exhaustivement le modèle de la sémantique des possibles argumentatifs exploité dans cette recherche. Ce modèle mixte de description-construction-représentation-interprétation de la signification lexicale, à partir d'informations linguistiques et discursives, s'inscrit dans la filiation directe de la sémantique argumentative, à laquelle il emprunte préceptes et notions ; une présentation préalable de la sémantique argumentative est donc effectuée.

Le chapitre 3, intitulé « *Principes de constitution du corpus TAL(N)* », vise un double objectif. Du point de vue théorique, il s'agit d'une part, en partant d'une définition courante de corpus, de parvenir à un concept scientifiquement exploitable de la notion de corpus électronique spécialisé, ce qui suppose parallèlement de choisir un cadre théorique approprié à la description d'une langue de spécialité, d'autre part de confronter et de sélectionner les différentes contraintes qui pèsent sur la constitution du corpus électronique spécialisé, compte tenu du cadre théorique adopté pour cette recherche. Du point de vue pratique, la section consacrée au corpus TAL(N) détaille les contraintes de constitution de corpus appliquées à la communauté TAL pressentie, puis la méthodologie d'obtention du corpus et son résultat.

Le chapitre 4, intitulé « *Hypothèses de recherche et protocole de validation* », étaye la problématique de la motivation argumentative en la dotant d'un ensemble d'hypothèses et d'une méthodologie adaptée à sa validation. Composé de deux parties, il expose d'abord le paradigme d'hypothèses constitué à partir du rapprochement des caractéristiques inhérentes

aux collocations textuelles et aux possibles argumentatifs ; ces hypothèses sont introduites dans un ordre allant du général au particulier. Il détaille ensuite le protocole de validation propre à tester ces hypothèses. Les ajustements pratiqués au niveau méthodologique se sont traduits par la création d'un processus de validation pluridisciplinaire mêlant linguistique de corpus et lexicographie, adapté à la nature électronique des données et aux contraintes lexicales liées à la sémantique des possibles argumentatifs.

Le chapitre 5, intitulé « *Analyse sémantico-discursive des collocations textuelles* », présente l'ensemble des analyses réalisées sur les six unités lexicales sélectionnées pour former les bases des collocations étudiées. Le chapitre est construit autour de trois grandes sections rassemblant les monovalents positifs *amélioration* et *performance*, les monovalents négatifs *erreur* et *problème* et les bivalents *évaluation* et *résultat*. Chacune de ces sections reprend ensuite la progression établie par la méthodologie et le processus de validation : dépouillement des données, construction de la signification lexicale et analyse des données en fonction de la nature grammaticale du collocatif. Le chapitre s'achève sur la synthèse de l'ensemble des régularités observées.

Chapitre 1

Collocations : définitions et problématiques

Présentation

« On dit *plonger dans le désespoir*, mais *mettre en rage* <**mettre en désespoir*, **plonger dans la rage*>. On peut être *gravement* ou *grièvement blessé*, mais seulement *gravement malade* <**grièvement malade*> (le fameux exemple de Bally) ; de plus, on a de *grands blessés*, alors qu'une blessure est *grave*, pas **grande*. On est *très fatigué*, mais *bien reposé* <**très reposé* [comme dans **Je suis très reposé*]>, *fabuleusement riche*, mais *bien connu* <**bien riche*, **fabuleusement connu*> ; et ainsi de suite. En français, on *fait* un pas, en espagnol, on le *donne* (= *dar un paso*) et en anglais, on le *prend* (= *take a step*). La pluie est *forte* en français, mais *lourde* (= *heavy rain*) en anglais. On *fait* un rêve en français, mais on l'*a* en anglais (= *have a dream*) ; vous *donnez* un cours en français, mais vous l'*enseignez* (= *teach a course*) en anglais et vous le *lisez* (= *citát' kurs*) en russe (même si vous ne le lisez pas !). Toutes ces bizarreries d'usage sont connues depuis longtemps : ce sont des **collocations** – des expressions phraséologiques d'un certain type, qu'on ne peut pas prévoir et que la linguistique ne savait pas présenter de façon logique et facilement calculable. » [Mel'cuk 1997, p. 23].

1 Les phénomènes phraséologiques

1.1 Définition de la phraséologie

Conformément aux propos de Mel’cuk ci-dessus cités, les collocations sont des expressions appartenant à la catégorie des phénomènes phraséologiques. La phraséologie est un vaste domaine d’investigations linguistiques¹, que Rey et Chantreau, dans leur *Dictionnaire des expressions et locutions*, définissent comme un « système de particularités expressives liées aux conditions sociales dans lesquelles la langue est actualisée, c’est-à-dire à des usages » [Rey et Chantreau 1989, préface p. IX] et que le TLFi décrit comme recouvrant l’« ensemble des tournures typiques d’une langue, soit par leur fréquence, soit par leur caractère idiomatique » [TLFi]. *Particularités expressives, usages, tournures typiques, caractère idiomatique...*, envisagée au sens lexicographique la phraséologie est un phénomène appartenant au patrimoine d’une langue [Mortchev-Bouveret 2004]..., envisagée au sens large [Cowie 1998] la phraséologie regroupe trois catégories de phénomènes : les expressions entièrement lexicalisées (ou figées), les expressions partiellement lexicalisées (ou figées) et les restrictions de sélection (syntaxiques et sémantiques), que Cruse appelle les *affinités* entre mots [Cruse 1986, p.18]. Les phénomènes phraséologiques rassemblent une telle variété de séquences linguistiques que la plupart des courants linguistiques (russe², anglais [Cowie 1998] ou français [Gonzalez Rey 2002]) ont adopté un principe de description fondé sur l’élaboration de typologies.

1.2 Présentation des phénomènes

La typologie des principales séquences phraséologiques proposée ci-dessous est empruntée à Grossmann et Tutin [Grossmann et Tutin 2003, pp. 6-8]. Ils distinguent tout d’abord les *séquences lexicales phrastiques* (ou propositionnelles) des *séquences lexicales syntagmatiques*.

Les séquences lexicales phrastiques correspondent à des expressions entièrement lexicalisées. Totalement figées, elles n’admettent aucune substitution possible. D’une façon générale, elles comprennent :

- les proverbes – **Ex** : *Pierre qui roule n’amasse pas mousse* ;
- les lieux communs – **Ex** : *À la guerre comme à la guerre* ;
- les maximes – **Ex** : *Tu ne tueras point* ;
- les slogans – **Ex** : *Tous pour un, un pour tous* !
- les formules usuelles contextuelles (appelées pragmatèmes par Mel’cuk et holophrases par Galatanu) – **Ex** : *Ça va ?*

¹ En témoigne la densité du programme proposé au colloque Phraséologie 2005, la phraséologie dans tous ses états, université catholique de Louvain (Belgique), 13-15 octobre 2005. <http://cecl.fltr.ucl.ac.be/PHRASEO/FRENCH/programmefr.html>

² Les phénomènes collocationnels ont également été traités (définitions et typologies) dans le cadre d’une tradition russe [Telija 1992, 1994, 1998 ; Baranov et Dobrovol’skii 1996].

Sur le plan sémantique, les expressions figées se caractérisent d'abord par leur non-compositionnalité sémantique, c'est-à-dire que le sens du tout ne peut se calculer à partir de la signification des parties. Chaque unité lexicale perd donc sa capacité référentielle. Ensuite, en particulier pour les expressions figées nominales, le référent est souvent unique³.

Ex : *nid d'ange*

Dans ce cas, le référent de l'expression ne renvoie ni à celui de *nid* ni à celui d'*ange*, même si on devine le procédé métaphorique sous-jacent, et l'expression désigne un vêtement pour enfant.

Sur le plan syntaxique, « la contiguïté des éléments, la non-insertion de modificateurs, la non-substituabilité des éléments caractérisent un sous-ensemble de ces expressions, mais, comme les collocations, ces paramètres ne sont pas définitoires de la classe » [Tutin et Grossmann 2002, p. 18].

Les séquences lexicales syntagmatiques sont des expressions partiellement lexicalisées (ou semi-figées). On peut y substituer un mot ou un syntagme (nominal, verbal, adverbial, adjectival). On y distingue les *unités phraséologiques* des *groupements usuels* (pour reprendre la terminologie introduite par Bally [Bally 1909]) par leur degré de figement.

Les unités phraséologiques désignent des groupes de mots où « les mots qui le(s) composent perdent toute signification et [...] l'ensemble seul en a une ; il faut en outre que cette signification soit nouvelle et n'équivalise pas simplement à la somme des significations des éléments (ce qui du reste serait absurde) » [Bally 1909, p.74]⁴. Selon l'opacité de la combinaison, on divise encore les unités phraséologiques en deux sous-catégories :

- les *locutions figées opaques* : le sens de l'expression n'est pas déductible de la signification des parties.

Ex : *cordon-bleu*

Un *cordon bleu* n'est ni un *cordon* ni *bleu*, le sens n'est pas compositionnel ;

- les *locution figées imagées* : le sens est imprédictible mais comporte une métaphore ou une métonymie perceptible.

Ex : *pomme de terre ; manger les pissenlits par la racine*

Le sens de ces expressions est non pas compositionnel mais partiellement dérivé de la signification première des parties.

³ Tutin et Grossmann [Tutin et Grossmann 2002] relativisent le caractère discriminant de ce critère comparé aux collocations, dans la mesure où certains collocatifs ont une fonction plus « typante » que qualifiante. L'exemple pris est celui de *café noir*, qui désigne à la fois une propriété du *café* et une sorte de *café*. Dans le premier cas, le référent est bien unique, mais pas tout à fait dans le second.

⁴ Les variations terminologiques pour décrire ces phénomènes sont nombreuses, on trouve également expressions figées, locutions, phrasèmes complets [Mel'cuk 1998], unités polylexématiques [Corbin 1997]...

Les groupements usuels ou *collocations* correspondent à des restrictions de sélection syntaxico-sémantique et se présentent comme « des expressions semi-figées binaires, constituées de deux expressions linguistiques »⁵ [Grossmann et Tutin 2003, p.8], la *base* et le *collocatif* pour reprendre la terminologie d'Hausmann [Hausmann 1989, 1997].

2 Les phénomènes collocationnels

2.1 Problèmes de définition

2.1.1 Une nature dynamique

À l'interface entre langue et discours, le concept de collocation est difficile à définir parce qu'il est dynamique. D'un point de vue quantitatif, les collocations sont importantes, car omniprésentes, tant en langue générale qu'en langue spécialisée, toutes langues confondues. En effet, une demi-page du *Nouvel Observateur* contient par exemple 21 collocations [Mel'cuk 2003, p. 26] et le *Dictionnaire spécialisé au français des affaires* [Binon et al. 2000] recense environ 3 000 vocables autour desquels gravitent pas moins de 11 000 collocations. D'un point de vue qualitatif, les collocations appartiennent bien au système de la langue, en tant que groupements usuels, syntaxiquement et sémantiquement contraints, mais leur développement change avec l'usage. Elles sont soumises à des contraintes sociolinguistiques de registre de langue : familier (*péter un boulon*, *péter la gueule à quelqu'un*), soutenu (*hanter la conscience*, *satisfaire un besoin*)... Elles témoignent également de l'appartenance d'un texte à un genre particulier (journalistique, *être en état d'alerte*, scientifique, *construire une hypothèse* [Pecman 2004a], littéraire, *clorre un chapitre*) et/ou à un domaine de pratique (affaires, *contrat de travail* [Binon et al. 2000], économie, *émettre des actions* [Cohen 1986], Internet, *visiter un portail* [Meynard 2000], pharmacologie, *dose d'attaque* [Laporte et L'homme 1997]). Enfin, elles subissent également les phénomènes de mode ; toutes les *questions* sont aujourd'hui *cruciales* [Grossmann et Tutin 2003, p. 6].

Le langage humain est un processus et non un produit ; aussi, ce qui est habituel évolue au cours du temps selon des facteurs régionaux, techniques ou sociaux. De nouvelles restrictions apparaissent tous les jours ; en ce sens, « la collocation doit donc être conçue comme un phénomène dynamique dont la signification est négociée et non figée » [Williams 2001b, p. 3].

2.1.2 Un continuum libre – figé

L'idée qu'il existe un stade intermédiaire de figement entre combinaisons libres et expressions figées n'est pas nouvelle. Elle apparaît déjà chez Bally, sous le terme de *séries phraséologiques* [Bally 1909, p. 66]. De fait, les collocations s'intègrent dans une zone qui oscille entre le syntagme libre et le syntagme figé, ce qui génère un triple problème de définition.

⁵ Dans le cas où une collocation associe plus de deux mots ou expressions, il s'agit généralement d'une superposition de collocations : *avoir une faim de loup* se décompose en deux collocations, *avoir faim* et *une faim de loup*.

Premièrement, les phénomènes collocationnels correspondent à des syntagmes semi-figés, se situant sur un continuum qui part du syntagme libre et qui va jusqu'au syntagme figé et dont il est difficile, voire impossible, de fixer les bornes. Ils sont situés dans une zone floue, que Thierry Fontenelle [Fontenelle 1997, p. 45] appelle le *fuzzy area*, entre la combinaison libre et l'expression figée, dont les frontières sont loin d'être étanches ; comment savoir où commence et où finit le semi-figé ?

Libre (*envie de chocolat*) → Semi-figé (*ferme intention*) → Figé (*pomme de terre*)

Deuxièmement, sous couvert du terme *collocation*, les phénomènes rassemblés sont de nature différente, de telle sorte qu'à l'intérieur de cette zone floue les restrictions syntagmatiques se distinguent également par différents degrés de figement. Sur un continuum allant du libre au figé⁶, les collocations sont alors qualifiées de régulière, transparente et opaque⁷.

Régulière (*colère destructrice*) → Transparente (*rouge de colère*) → Opaque (*colère noire*)

Les limites entre les différentes catégories de collocations ne sont pas toujours faciles à établir, notamment à l'intérieur des collocations régulières. Certaines collocations s'associant parfois davantage à des combinaisons libres, elles s'approchent ainsi sensiblement de ce que l'on pourrait appeler à l'instar de Gross [Gross 1995], des classes d'objets : « Les domaines d'arguments de prédicats (verbaux, nominaux ou adjectivaux), quand ils sont libres, peuvent être exprimés en termes de classes d'objets [...] et non d'unités lexicales. Le verbe ressembler a comme objet un élément de la classe des chaussures (soulier, sandale, botte, etc.). Le verbe lire est caractérisé par deux types de classes : celle des textes (récit, dissertation, roman, rédaction, poème, article, etc.) et celle des supports de textes (livres, journal, cahier, brochure, etc.) ». [Gross, 1996, pp. 78-79].

2.1.3 Une approche prototypique

Les collocations sont également des phénomènes intuitifs pour tout locuteur natif, pouvant être décrits en termes de prototypes⁸. En effet, à la manière de Martin [Martin 1997b] pour les locutions, il est possible de citer des cas typiques qui fassent l'unanimité et donc tentant de les utiliser comme supports permettant de couvrir les multiples facettes – pas toujours convergentes – du concept de *collocation*.

Ex : *gravement malade ; grièvement blessé ; désirer ardemment ; aimer éperdument*

Les célèbres exemples de Bally [Bally 1909] illustrent clairement cette dimension prototypique.

⁶ Une analyse diachronique montre une tendance des collocations à évoluer du statut de collocation régulière à celui de collocation opaque, ce qui corrobore leur aspect dynamique.

⁷ Nous recourrons à la typologie sémantique proposée par Tutin et Grossmann [Tutin et Grossmann 2002] (cf. Section 2.3.1.2.1).

⁸ À noter que Tutin et Grossmann [Tutin et Grossmann 2002] évoquent le problème en n'envisageant qu'une catégorie prototypique possible, tandis que Williams [Williams 2001] en projette plusieurs selon le degré de figement de la restriction syntagmatique considérée, à l'intérieur de la classe formée par l'ensemble des phénomènes collocationnels.

Toutefois, deux raisons poussent Tutin et Grossmann à nous mettre en garde contre cette tentation. D'une part, « une approche par prototype, pour intuitive qu'elle soit, n'apparaît en effet pas suffisamment contraignante pour avoir une capacité prédictive ou explicative » [Grossmann et Tutin 2003, p. 9]. D'autre part, le choix d'un prototype ne ferait, selon eux, « que reporter l'étude systématique des critères linguistiques pourtant indispensable pour des traitements formels » [Tutin et Grossmann 2002, p. 5].

À défaut de pouvoir définir objectivement et simplement les phénomènes collocationnels, les linguistes cherchent dans un premier temps à décrire ce qui peut être accepté en qualité de collocation et établissent à cet effet des critères.

2.2 Solutions de description

2.2.1 Critères de figement

La comparaison des critères de description proposés par Williams [Williams 2001b] et par Tutin et Grossmann [Tutin et Grossmann 2002] montre que seuls deux paramètres plaçant les phénomènes collocationnels sur une échelle de figement concordent, bien qu'ils soient discutés. Ces paramètres correspondent à leur caractère transparent et à leur aspect arbitraire.

2.2.1.1 Critère de la transparence

Le critère de la transparence ne s'applique que relativement aux collocations lexicales, et non aux collocations grammaticales comme *se souvenir de*, eu égard à la dimension sémantique qu'il mobilise. En effet, il réfère à leur caractère interprétable malgré leur imprévisibilité (*cf.* Section 2.2.1.2 sur le critère de l'arbitraire) : « The term collocation will be used to refer to sequences of lexical items which habitually co-occur, but which are nonetheless fully transparent in the sense that each lexical constituent is also a semantic constituent. » [Cruse 1986]. De fait, n'importe quel locuteur peut en déduire le sens, alors même que, dans le cas d'un locuteur non natif, les collocations sont impossibles à produire en l'absence de mémorisation préalable.

Ex : *buveur invétéré ; heavy drinker*

Toutefois, ce critère semble échouer à caractériser l'ensemble de la classe collocationnelle puisque, parfois, le sens ne semble pas réellement transparent.

Ex : *colère noire ; peur bleue*

Dans ces cas, le sens du collocatif n'est pas immédiatement interprétable, le caractère transparent est donc discutable.

2.2.1.2 Critère de l'arbitraire

Les collocations sont également par nature arbitraires : « Collocations should be defined not just as “recurrent word combinations”, but as “arbitrary word combinations”. » [Benson 1989, p. 3] (*cf.* Section 2.2.2.2 sur le critère de récurrence). Le critère arbitraire qualifie la non-prédictibilité de l'association lexicale : « Si *torrentielle* peut en effet apparaître en cooccurrence avec *pluie*, cela paraît nettement moins naturel avec *précipitations torrentielles*.

Cela est encore plus manifeste pour les collocations imagées du type *appétit d'ogre* ou *faim de loup* (vs. ?*appétit de loup* et ?*faim d'ogre*)⁹. » [Tutin et Grossmann 2002, p. 9]. Cet aspect arbitraire est doublement significatif dès lors que l'on aborde le problème de la traduction des collocations. D'une part, d'une langue à l'autre les bases ne sélectionnent pas les mêmes collocatifs :

Ex : *circulation intense* se traduit en anglais par *heavy traffic*, non par **circulation lourde*. D'autre part, il faut distinguer entre ce qui peut être librement traduit et ce qui exige la connaissance collocationnelle, sachant que ce qui est non arbitraire (un syntagme libre) dans une langue peut ne pas l'être dans une autre [Williams 2001b].

Par ailleurs, le caractère non prédictible de l'association lexicale caractérise principalement l'aspect sémantique qui unit le collocatif à la base, car du point de vue syntaxique les collocations suivent des patrons précis. « Par exemple, les collocations construites autour d'un nom incluront des adjectifs ou des verbes, mais probablement pas des adverbes ou des conjonctions. » [Tutin et Grossmann 2002, p. 9]. De même que pour le critère de transparence, ce critère nécessite d'être relativisé, au risque d'échouer à caractériser l'ensemble de la classe collocationnelle puisque, parfois, l'association lexicale ne semble pas réellement arbitraire.

Ex : *l'âne brait*

Du point de vue sémantique, l'aspect arbitraire de l'association est discutable, et la collocation suit un patron syntaxique connu.

2.2.2 Critères définitoires

2.2.2.1 Critères de Tutin et Grossmann

Les critères définitoires sont qualifiés par Tutin et Grossmann [Tutin et Grossmann 2002] de propriétés équivalant à des conditions nécessaires. Ils sont au nombre de trois : le caractère binaire de la collocation, la dissymétrie de ses composants et sa cooccurrence restreinte.

2.2.2.1.1 Le caractère binaire

La communauté scientifique s'accorde généralement à décrire une collocation comme étant composée de deux mots ou de deux lexies¹⁰. Mel'cuk utilise également le terme d'« expression linguistique bipartite AB » [Mel'cuk 2003, p. 23].

Dans le cas où la collocation englobe plus de deux mots, elle se laisse généralement décomposer en deux collocations distinctes.

Ex : *essuyer un échec cuisant* se laisse décomposer en *essuyer un échec* et *échec cuisant*

Toutefois, la fusion des collocations n'est pas toujours réelle.

Ex : *prendre peur* et *peur bleue* ne permettent pas de générer **prendre une peur bleue*

⁹ Conformément aux pratiques linguistiques traditionnelles, les représentations problématiques sont notées « ? » et les représentations asémantiques ou agrammaticales, « * ».

¹⁰ Cette bipartition n'est pas aussi tranchée chez Sinclair, « Collocation is the occurrence of two or more words within a short space of each other » [Sinclair 1991, p. 170], et Cruse [Cruse 1986], par exemple.

Dans le cas où la collocation est composée d'un syntagme, Tutin et Grossmann [Tutin et Grossmann 2002, p. 10] préconisent de parler « d'association de constituants » plutôt que de mots ou de lexies.

Ex : *fort comme un turc* : *fort* est un mot et *comme un turc* est un syntagme

2.2.2.1.2 La dissymétrie des composants

Pour Hausmann et Mel'cuk, les deux composants d'une collocation n'ont pas le même statut : « Dans la collocation, le statut des deux partenaires combinés n'est pas égal. » [Hausmann 1989, cité dans Tutin et Grossmann 2002, p. 10]. Pour Hausmann, la base est autonome, elle conserve sa signification, tandis que le collocatif est assujéti à la base. De même, pour Mel'cuk, le rapport entre les deux éléments d'une collocation est dissymétrique, puisque la base conserve toujours sa signification habituelle.

Ex : *peur bleue*

Dans cette collocation, le nom *peur* conserve sa signification habituelle – telle qu'elle est susceptible de figurer dans un dictionnaire de langue française –, tandis que la signification de l'adjectif *bleu* est quelque peu détournée, au profit d'un sens particulier à ce rapprochement lexical.

2.2.2.1.3 La cooccurrence restreinte

Le critère concernant la cooccurrence restreinte apparaît comme la continuité du caractère dissymétrique précédemment exposé. En effet, le premier critère témoigne de la faculté de la base à conserver sa signification première, tandis que ce second critère montre la propension du deuxième élément (ou collocatif) à dépendre de la base de façon irrégulière et/ou contrainte, pour reprendre la terminologie de Mel'cuk, auquel nous empruntons les deux exemples ci-dessous [Mel'cuk 2003, p. 24].

Ex : *café noir*

En français, un *café noir* est un café – sans produit laitier –. Si *café* est choisi par le locuteur sans restriction et conserve de fait sa signification habituelle, non seulement le locuteur n'a pas le choix du mot à utiliser pour exprimer ce concept, mais en plus la signification de l'adjectif *noir* est quelque peu détournée. Sous l'entrée *noir*, aucun dictionnaire ne mentionne la signification « sans produit laitier », parce que cet adjectif ne prend ce sens que combiné avec *café*.

Ex : *année bissextile*

Dans la collocation *année bissextile*, – année qui a 366 jours –, l'adjectif *bissextile* a la signification de – qui a 366 jours –, mais il ne se combine qu'avec *année* (**an bissextil*). Il est donc sélectionné à cause de sa signification extrêmement précise et de façon contrainte.

2.2.2.2 Critères de Williams

Les critères discutés par Williams [Williams 2001b] sont au nombre de deux : le caractère habituel de la collocation et sa bonne formation syntaxique.

2.2.2.2.1 *Le caractère habituel*

Historiquement, le caractère habituel figure parmi les premiers critères énoncés pour la reconnaissance des collocations : « Collocations of a given word are statements of the habitual or customary places of that word in collocational order. » [Firth 1957, cité dans Williams 2001b, p. 2]. La difficulté posée par ce critère provient de ce qu'il est effectivement *habituel* de rencontrer de tels binômes, sans pouvoir instancier la réalité qu'*habituel* recouvre. Se contenter de ce « qui est (devenu) fréquent, régulier, coutumier » [TLFi] n'est pas satisfaisant, au sens où une telle définition ne permet pas de disposer de paramètres quantitativement discriminants, comparés aux cooccurrences de syntagmes libres. C'est pourquoi est généralement associée au critère habituel une dimension statistique, ce que Sinclair formule ainsi : « SIGNIFICANT COLLOCATION is regular collocation between two items, such that they co-occur more often than their respective frequencies and length of the text in which they appear would predict. » [Sinclair et al. 1970, p. 150]¹¹. Cette dimension statistique prend généralement la forme d'une mesure, les plus courantes étant le z-score, le t-score et l'information mutuelle, que nous détaillons ultérieurement (*cf.* Section 2.3.2.3.2 sur la typologie statistique).

La limite posée par ce critère tient au fait que chaque mesure statistique est appropriée pour une application donnée, ce paramètre définitoire du concept *collocation* demeure donc fluctuant, puisqu'il est fonction du calcul appliqué.

2.2.2.2.2 *La bonne formation syntaxique*

Le critère de grammaticalité tel qu'il est présenté par Williams [Williams 2001b] n'est envisagé que dans la perspective d'une extraction automatique des collocations. De ce fait, ce critère est légitime d'un point de vue lexicographique pour limiter le bruit [Kjellmer 1984] ou, selon les besoins de l'étude, ne sélectionner que les collocations lexicales, comme l'outil Xtract le fait par ajout de filtres linguistiques [Smadja 1993]. Toutefois, Williams fait remarquer que « les collocations textuelles conformes à la définition de cooccurrence dans une fenêtre utilisée par Sinclair n'ont pas besoin de ce critère puisqu'elles recherchent des associations dans un champ sémantique » [Williams 2003, p. 5]. Firth précise lui-même que les « collocations of a given word are statement of the habitual or customary places of that word in collocational order but not in any other contextual order and emphatically not in any grammatical order » [Firth 1957, cité dans Williams 2003, p. 5]. Plusieurs recherches sur les thèmes [Berry-Roghe 1973 ; Phillips 1985], sur la polysémie [Clear 1994] ou sur les réseaux collocationnels [Williams 2001b] se dispensent du critère de grammaticalité.

¹¹ À noter l'évolution du point de vue de Sinclair entre 1970 (two items) et 1991 (two or more words), concernant le caractère binaire des collocations.

Les critères élaborés et appliqués sont fonction du cadre théorique et du domaine de pratique dans lesquels la recherche se situe. Par répercussion, les conditions nécessaires retenues influent sur les paramètres définitoires entrant dans la définition de collocation mobilisée, ce qui engendre une variation des définitions du concept selon les approches et les problématiques abordées.

2.3 Deux approches du concept de « collocation »

2.3.1 Approche lexicologique – lexicographique

2.3.1.1 Définition restreinte

D'une façon générale, dans la tradition lexicologique [Cruse 1986] et lexicographique [Hausmann 1989 ; Mel'cuk 1998], une conception « restreinte » de la collocation prédomine, entendue comme une association lexicale syntagmatique restreinte entre deux éléments entretenant une relation syntaxique. Cette définition prend sa source dans des études théoriques qui se caractérisent par la volonté d'aboutir à une définition formelle du concept. Reformulant à leur façon la formalisation élaborée par Mel'cuk [Mel'cuk 1998], Tutin et Grossmann proposent la définition suivante : « Une **collocation** est l'association d'une **lexie L (mot simple ou phrasème¹²)** et d'un **constituant C** (généralement une lexie, mais parfois un syntagme, par exemple « à couper au couteau » dans *un brouillard à couper au couteau*) entretenant une relation telle que :

- C (le collocatif) est sélectionné en production pour exprimer un sens donné en cooccurrence avec L (la base) ;
- le sens L est habituel. » [Tutin et Grossmann 2002, p. 11].

Cette définition met en avant les paramètres sémantiques et syntaxiques inhérents à l'association collocationnelle, sans résoudre le problème de la diversité des éléments convoqués, auxquels il n'est pas pertinent d'appliquer un traitement homogène. C'est pourquoi, en complément de principes définitoires, les chercheurs s'appuient généralement sur des typologies classant les combinaisons lexicales possibles en fonction des paramètres sémantiques et/ou syntaxiques.

2.3.1.2 Typologies

2.3.1.2.1 Typologie selon des critères sémantiques

Tutin et Grossmann [Tutin et Grossmann 2002] proposent une « ébauche » de typologie des collocations fondée sur des critères sémantiques prenant en compte l'idiosyncrasie et l'idiomaticité¹³ du collocatif. Ils distinguent trois sortes de collocations :

¹² Expression complètement lexicalisée dans la terminologie mel'cukienne.

¹³ L'idiomaticité est liée au caractère non standard de l'association linguistique, tandis que l'idiosyncrasie est liée au caractère unique de cette association [Grossmann et Tutin 2003, p. 9].

- **les collocations opaques.** C a un sens particulier en cooccurrence avec L différent de sa signification en dehors de cette association. La base conserve sa signification habituelle. Le collocatif semble imprédictible et l'association arbitraire.

Ex : *nuit blanche* ; *peur bleue*

Les collocations opaques sont proches des expressions figées (comme *cordon-bleu*), à la différence près que la base reste interprétable. Les collocatifs sont souvent peu productifs ou donnent lieu à des séries limitées (comme *bleu* avec *frousse* ou *trouille*) ;

- **les collocations transparentes.** C semble conserver sa signification habituelle mais être difficilement prédictible du point de vue sémantique et/ou syntaxique.

Ex : *désir ardent*/?*désir enflammé* ; imprédictibilité sémantique (quasi-synonymie des adverbes sans substitution possible)

Ex : *prendre peur* ; imprédictibilité syntaxique (absence d'article, **prendre la peur*)

Les collocations transparentes constituent le cas prototypique. Le décodage est rendu possible par un sens du collocatif proche de sa signification ou une périphrase comportant une comparaison ou une métaphore facilement interprétable ;

- **les collocations régulières.** C inclut la signification de L ou a une signification très générique. L'association est motivée et transparente. Soit le collocatif est à base unique, soit il porte sur un paradigme qui semble pouvoir être défini par des traits sémantiques.

Ex : *nez aquilin* ; base unique (*aquilin* se définit à l'aide de la base *nez*)

Ex : *grande tristesse* ; paradigme sémantique (trait intensif porté par *grande*)

Les collocations régulières sont proches des expressions libres (comme *réelle tristesse*).

La conception étroite de la définition de collocation s'accompagne de restrictions syntaxiques dans la combinatoire des cooccurrents, soit envisagées en combinaison avec une typologie fondée sur des critères sémantiques¹⁴, soit envisagées seules.

2.3.1.2.2 Typologie selon des critères syntaxiques

L'association d'une base et d'un collocatif génère un syntagme qui mobilise des constituants spécifiques. L'association de ces constituants n'est pas arbitraire et permet de générer des syntagmes nominaux (nom + adjectif) ou verbaux (verbe + adverbe). C'est pourquoi Hausmann [Hausmann 1989] décrit la classe des collocations par une typologie des paires de constituants que les collocations peuvent générer. Suivant cette démarche, une

¹⁴ Il apparaît par exemple que les collocations régulières sont les moins contraintes syntaxiquement.

collocation correspond à la combinaison caractéristique de deux mots dans une des structures suivantes¹⁵ :

- **nom + adjectif (épithète)** : *amour platonique* ;
- **nom + (prép.) + nom** : *bourreau des cœurs* ;
- **nom (sujet) + verbe** : *la colère s'apaise* ;
- **verbe + nom (objet)** : *commettre une agression* ;
- **verbe + (prép.) + nom** : *rougir de honte* (Hausmann 1999) ;
- **verbe + adverbe** : *exploiter efficacement* ;
- **adverbe + adjectif** : *sexuellement transmissible*.

Toutefois, Tutin et Grossmann [Tutin et Grossmann 2002, p. 13] font remarquer que la liste des patrons syntaxiques donnée par Hausmann n'est pas exhaustive, comparée, entre autres, aux collocations présentes dans le *Dictionnaire explicatif et combinatoire* – DEC – [Mel'cuk et al. 1984, 1988, 1992, 1999]. Par exemple, les syntagmes construits sur des bases nominales admettent non pas uniquement des adjectifs mais également des prépositions comme *de joie*. De même, les syntagmes construits sur des bases adjectivales admettent non pas uniquement des adverbes mais aussi des syntagmes prépositionnels comme *ivre de colère* ou d'autres adjectifs comme *ivre mort*.

Les limites posées par ce type de typologie proviennent du fait que les patrons syntaxiques des collocations sont susceptibles de variations. Par exemple, la contiguïté de la base et du collocatif n'est pas systématique. On peut dire *j'aime le steak bleu*, mais également *mon steak est vraiment bleu*. De même, certaines collocations verbales autorisent des alternances syntaxiques. On peut dire *mener une attaque*, mais aussi *l'attaque a été menée* (passivation). Constatant ces limites, Tutin et Grossmann [Tutin et Grossmann 2002] émettent un doute quant à la capacité d'un modèle fondé sur une typologie des constituants syntaxiques à rendre compte de ces associations.

¹⁵ La typologie d'Hausmann repose sur des éléments définitoires identiques à ceux évoqués préalablement (section 2.3.1.1). Les collocations sont des combinaisons de mots binaires aux capacités combinatoires limitées, dont les composants (la base et le collocatif) entretiennent une relation hiérarchisée – la base est sémantiquement autonome et conserve son sens habituel –, tandis que le collocatif est sélectionné en fonction de la base pour exprimer un sens donné [Hausmann 1989, 1997].

2.3.2 Approche « linguistique de corpus »

2.3.2.1 Historique

2.3.2.1.1 *L'origine contextualiste*

En réaction contre les courants linguistiques structuralistes, qui évacuent la question du sens des mots, l'approche pragmatique prônée par Firth place l'étude de la langue en contexte. Influencé par les travaux anthropologiques de Malinowski, qui écrit que « the real knowledge of a word comes through the practice of appropriately using it within a certain situation » [Malinowski 1923, p. 325], Firth reprend l'idée de contexte « afin d'insister sur la nécessité d'exploiter ce qui est visible plutôt que d'élaborer des hypothèses abstraites sur la notion de sens » [Williams 2003, p. 34]. Dans cette perspective, l'analyse des collocations participe de l'élaboration d'une théorie du sens contextualiste. Aussi, lorsque Firth [Firth 1957] introduit pour la première fois le terme collocation, il entend caractériser certains phénomènes linguistiques de cooccurrence qui engendrent des modifications de sens constatées sur les mots par leur relation de collocation en contexte : « Meaning by collocation is an abstraction at the syntagmatic level and is not directly concerned with the conceptual or idea approach to the meaning of words. One of the meanings of night is its collocability with dark, and of dark, of course, collocation with night. » [Firth 1957, p. 196]. Le contexte est accessible par le texte, donc, concrètement, l'analyse des collocations se fonde sur l'observation des relations sémantiques créées par le contexte, à la surface des textes.

Dans la lignée contextualiste, plusieurs recherches ont été menées en vue de mettre en pratique les principes théoriques de Firth. Halliday s'est attaché à poursuivre la description du système collocationnel en préservant l'importance de l'environnement textuel : « Without our being aware of it, each occurrence of a lexical item carries with it its own textual history, a particular collocational environment that has been built up in the course of the creation of the text and that will provide the context within which the item will be incarnated on its particular occasion. This environment determines the 'instantial meaning', or text meaning, of the item, a meaning which is unique to each specified instance. » [Halliday and Hasan 1976, p. 289]. Conformément aux postulats firthiens, le mot dérive son sens du contexte et influe simultanément sur ce contexte pour créer l'environnement textuel. La définition du texte est donc primordiale, puisque c'est sur lui que repose toute étude. Pour Halliday, le texte doit représenter une unité sémantique, un tout qui possède une cohésion. La collocation est un élément de cette cohésion, elle a un rôle fonctionnel de cohésion textuelle : « Laugh... joke, blade... sharp, ill... doctor [...] The cohesive effect of such pairs depends not so much on any systematic relationship as on their tendency to share the same lexical environment, to occur in COLLOCATION with one other. In general, any two lexical items having similar patterns of collocation – that is, tending to appear in similar contexts – will generate a cohesive force if they occur in adjacent sentences. » [Halliday and Hasan 1976, pp. 285-286]. La fonction de la cohésion textuelle, assurée par les liens lexicaux entre énoncés proches et distants, a été confirmée par Hoey, dont l'origine des travaux repose sur la structuration et les stratégies rhétoriques des textes : « [...] Collocation is the direct product of bonding (and of similar relations between sentences in and across texts). Collocation does not exist in language independent of its grounding in text, and we acquire collocations, as acquire other aspects of language, through encountering texts in the course of our lives. » [Hoey 1991, p. 219].

2.3.2.1.2 *Les développements de l'école de Birmingham*

À partir des années 1960, le centre de linguistique contextualiste s'est déplacé à Birmingham, sous l'égide de Sinclair [Sinclair 1991]. Le lancement du projet OSTI [Sinclair et al. 1970, 2002] marque alors un tournant dans les recherches sur l'émergence du sens en contexte. En effet, ce rapport démontre la nécessité de travailler sur de grands ensembles de textes, les seuls à pouvoir révéler la fonction cohésive des phénomènes collocationnels en contexte. Ce rapport eut pour effet l'élaboration d'un autre projet : le COBUILD, premier grand corpus électronique constitué à des fins lexicographiques, dans lequel l'environnement contextuel des mots devient une réalité descriptible. La nature électronique du corpus ainsi constitué, le développement d'outils informatiques adaptés à l'exploration de ces données textuelles et les analyses qui en résultent ont été à la base d'une nouvelle discipline : la linguistique de corpus. Aussi, le véritable essor de l'école contextualiste est venu avec le développement de cette discipline dans les années 1980, puisque la linguistique de corpus offre des possibilités techniques nouvelles pour analyser de grandes quantités de données à partir desquelles peuvent émerger des régularités parmi lesquelles se trouvent les collocations. Cette approche empirique a permis de remplacer l'introspection fondée sur l'intuition du chercheur par une approche inductive : les hypothèses formulées s'appuient sur des données réelles et le corpus est considéré comme révélateur de la langue [Tognini-Bonelli 2001, p. 3].

2.3.2.2 **Définition large**

Firth a introduit et décrit le concept de collocation par des exemples, sans toutefois en proposer de réelle définition ni de formalisation. Nous avons vu que les recherches visant à appliquer les principes du sens contextualiste ont abouti à la mise en évidence d'un principe de cohésion textuelle selon lequel les collocations jouent un rôle essentiel, principe à partir duquel a émergé la notion de collocation textuelle. La définition de collocation textuelle mobilise le sens premier du concept de collocation, c'est-à-dire qu'elle désigne des relations de sens portées par des mots qui tendent à apparaître ensemble dans un contexte textuel donné. Aussi, sur le plan sémantique, l'analyse des collocations se dispense de toute étude lexicale préalable. Le syntagme en tant que formulation quasi idiomatique est délaissé au profit du sens contextuel du mot. De même, les relations analysées étant d'ordre pragmatique, dans la perspective contextualiste et néo-contextualiste, le critère de bonne formation syntaxique n'a tout simplement pas lieu d'être : « Les collocations textuelles conformeront à la définition de cooccurrence dans une fenêtre utilisée par Sinclair n'ont pas besoin de ce critère puisqu'elles recherchent des associations dans un champ sémantique » [Williams 2001b, p. 13]. Cette conception large de collocation « ne restreint nullement l'association à la pure contiguïté et englobe aussi bien des associations lexicales paradigmatiques (médecin... hôpital) que des associations lexicales syntagmatiques (argument de poids), voire des associations incluant des mots grammaticaux (se souvenir de) » [Tutin et Grossmann 2002, p. 8]. Toutefois, Sinclair considère que seules les collocations lexicales méritent l'appellation de collocation : « Collocation in the purest sense ... recognises only the lexical co-occurrence of words. » [Sinclair 1991, p. 170]. D'ailleurs, ce principe était déjà sous-entendu chez Firth car, parallèlement au terme *collocation*, il introduit le terme « *colligation* » pour désigner les collocations grammaticales [Firth 1957, p. 190]. Cette distinction se retrouve également dans la typologie de Benson [Benson et al. 1986], qui définit la collocation grammaticale comme « a phrase consisting of a dominant word (noun, adjective, verb) and a preposition or grammatical structure such as an infinitive or clause » [Benson et al. 1986, préface p. IX].

Cette acception uniquement textuelle de collocation, cumulée aux développements de la linguistique de corpus, a eu pour effet d'éloigner progressivement les critères liés à la compétence des locuteurs natifs, liés aux aspects sémantiques (lexicaux) et syntaxiques, au profit d'une instanciation statistique de la collocation textuelle. Dans la lignée de Haskell [Haskell 1971], Jones et Sinclair [Jones and Sinclair 1974], Church et Hanks [Church and Hanks 1990], Aarts [Aarts 1990] et Sinclair [Sinclair 1991], la définition statistique de la collocation textuelle est la suivante : « A collocation is a group of two or more words which are to be found in proximity to each other significantly more often than one would predict, given the frequency of occurrence of each word taken individually. They may or may not be immediate neighbours, and the meaning of the whole may or may not be fully determined by the meaning of the parts. » [Kilgarriff 1992, pp. 29-30].

2.3.2.3 Typologies

2.3.2.3.1 Typologie selon des critères contextualistes

La typologie proposée par Partington [Partington 1998] est qualifiée de contextualiste, dans la mesure où les trois types de collocations proposés correspondent à « une interprétation différente de l'injonction de Firth selon laquelle “ you shall know a word by the company it keeps ” » [Williams 1999b, p. 134].

- **la collocation textuelle.** Elle se définit comme la cooccurrence de deux mots ou plus dans un espace de texte court ;
- **la collocation psychologique** (ou associative). Elle doit son existence aux compétences linguistiques permettant de statuer intuitivement sur ce qui est « normal » et/ou « correct » et désigne par voie de conséquence l'association de deux mots ou plus, qui ne semble transparente qu'à un locuteur natif ;
- **la collocation statistique.** Elle correspond à la probabilité de rencontrer deux mots ou plus juxtaposés dans un espace de texte court.

En l'absence de critère syntaxique ou sémantique propre, la collocation textuelle est également une collocation statistique [Williams 1999b, p. 145].

2.3.2.3.2 Typologie selon des critères statistiques

Les statistiques ne définissent pas les collocations mais sont un moyen d'évaluer leur seuil de signification, car « seules des mesures quantitatives peuvent révéler les faits de langage dans une masse de données » [Williams 2003, p. 37], d'où le terme de collocation significative utilisé par Sinclair [Sinclair et al. 1970]. Les analyses statistiques ont montré que les collocations sont significatives dans un espace de texte court qui correspond généralement à une fenêtre de + ou - 4 mots à partir du nœud. Les contextualistes parlent de nœud [Jones and Sinclair 1974], et non de base au sens d'élément stable dans une collocation comme Hausmann [Hausmann 1979], parce que le rôle sémantique joué par le nœud n'est pas pris en considération dans le calcul.

Il existe de nombreuses mesures d'association permettant d'identifier les unités lexicales qui coexistent le plus souvent dans une collection de textes. Les deux plus courants sont le t-score et l'information mutuelle – IM. L'algorithme t-score mesure le degré d'association entre

deux mots, en favorisant l'émergence des combinaisons dont la fréquence est élevée. L'algorithme IM « ...compares the probability of observing *x* and *y* together (the joint probability) with the probabilities of observing *x* and *y* independently » [Church and Hanks 1990, p. 23]. À l'inverse, l'IM élimine les fréquences importantes au profit de fréquences plus faibles ; c'est une mesure de confiance. Les deux algorithmes sont donc complémentaires, ils permettent l'émergence d'unités différentes. Le *t-score* « by identifying frequent and very reliable collocations, offers the lexicographer a semantic profile of the node word and a set of particular fixed phrases, grammatical frames and typical stereotyped combinations » [Clear 1993, p. 282], tandis que l'IM « is best consulted for information about pairs which, though not likely to be typical of the usage of the node word, will be strongly associated and tend to form idioms, proverbs and technical phrases » [Clear 1993, p. 282].

À l'image des autres critères (binaire, arbitraire, syntaxique...), le critère statistique a également ses limites [Wouden 1997]. La mesure *t-score* pose problème car, en sélectionnant des unités à haute fréquence, elle ramène beaucoup de combinaisons composées de mots grammaticaux. C'est pourquoi la liste des combinaisons recueillies correspond plus exactement à un ensemble de candidats collocations – de collocations potentielles –, qu'il convient de nettoyer pour en extraire les collocations pertinentes au vu des objectifs de la recherche en cours. L'IM présente plusieurs inconvénients. Outre le fait de ne pouvoir différencier les binômes à sémantique compositionnelle des autres et regrouper les mots par classe grammaticale, l'IM a tendance à privilégier les collocations atypiques et bizarres. Pour limiter cet effet indésirable, une solution consiste à éliminer les hapax et les unités de faible fréquence du champ de calcul. Toutefois, cette mesure a les défauts de ses qualités, puisque « la convergence de différentes expériences montre cependant qu'en donnant un poids important aux événements rares et en soulignant les emplois « spécialisés » le score de cooccurrence fait ressortir les expressions figées, ce qui est précieux dans une perspective lexicographique » [Habert et al. 1997, p. 118].

Parallèlement aux mesures statistiques, qui permettent de « calculer les collocations », certaines typologies permettent de les classer relativement à ce critère statistique. Sinclair définit par exemple les *collocations descendantes* (« When *a* is node and *b* is collocate, I shall call this *downward collocation* – collocation of *a* with a less frequent word *b*. » [Sinclair 1991, p. 115]) et les *collocations montantes* (« When *b* is node and *a* is collocate, I shall call this *upward collocation*. » [Sinclair 1991, p. 116]) en fonction de la fréquence des composants (le nœud et le collocatif) de la collocation. Nous reprenons ses exemples pour illustrer les deux concepts. Prenons la collocation anglaise *to give an audience*, et supposons que les mots *to give* et *audience* reviennent respectivement cinq cents fois et trente fois dans un corpus de texte. Pour Sinclair, la relation collocationnelle entre ces deux composants change en fonction du mot à l'étude. Si le lexicographe s'intéresse au verbe *to give* et trouve que ce mot est fortement associé à *audience*, cette combinaison sera descendante (*downward collocation*) puisque le mot *audience* est moins fréquent que le mot à l'étude. En revanche, si le lexicographe s'intéresse au mot *audience* plutôt qu'à *to give*, la collocation est qualifiée de montante (*upward collocation*), puisque le verbe *to give* est plus fréquent que le substantif *audience*.

La mesure statistique des collocations textuelles s'accompagne d'un processus d'extraction automatique des données. Les fonctions statistiques sont souvent incorporées dans des outils informatiques, pour l'extraction automatique comme Xtract [Smadja 1993] ou pour l'extraction terminologique comme ACABIT [Daille 1996a, 1996b ; Daille and Williams 2001].

2.3.2.3.3 Typologie selon des critères sémantiques

Une définition statistique de la collocation n'exclut pas la possibilité d'une classification indexée sur une typologie fondée sur des critères sémantiques. À la différence de celles fondées sur une définition restreinte de collocation (*cf.* Section 2.3.1.1), ces typologies sont plus larges, au sens où elles incluent généralement le concept de *collocation ouverte* ou *libre*. Par exemple, Cowie [Cowie 1978 ; Cowie et al. 1983] propose une typologie qui dissocie *open collocation* et *restricted collocation*. La première catégorie inclut des associations de mots dont les deux éléments conservent leur sens littéral et dont la combinatoire est assez libre pour sembler quasi illimitée.

Ex : le verbe *to run* se combine avec *machine, business, horse, program...*

La seconde catégorie inclut, de façon traditionnelle, des associations de mots dont l'un des constituants conserve son sens littéral tandis que l'autre a un sens figuré. Dans ce cas, le sens du mot A limite les possibilités combinatoires de l'élément B.

Ex : *to explore a claim ; to kill a cigarette*

De même, Howarth [Howarth 1998] propose une typologie qui répartit les phénomènes collocationnels en cinq catégories, selon leur degré progressif de restriction sémantique, allant des *free collocations* aux *complete restrictions*. La première catégorie : « Free collocations, for example, consist of two elements, both of which are used in their literal senses, and which permit the substitution of either element without affecting the sense of the other component. In the following examples both the verbs publish and put are used in their primary senses, and the substitution of other noun objects (such as *publish a book / an article / a prospectus* or *put a comment / an address / full stop*) makes no difference to the sense of the verb. » [Howarth 1998, p. 167]. Tandis que la dernière catégorie – *complete restrictions* –, rassemble des cas où, par exemple, aucun autre nom ne peut être utilisé avec tel verbe employé dans un sens donné.

Ex : *curry favour*

Une typologie aussi large couvre naturellement l'ensemble des phénomènes collocationnels possibles, la difficulté à laquelle Howarth n'échappe pas étant de trancher sur le statut sémantique des éléments ainsi rassemblés : « A decision can then be taken concerning where the cut-off point is between restricted and free collocations. According to the present analysis collocability at Level 1 is felt to be insufficiently constrained to be restricted, so Levels 2-5 constitute the category of restricted collocations. » [Howarth 1998, p. 168].

Les variations de définitions et la diversité des critères permettant de décrire les collocations sont issues des différentes méthodes de traitement et d'analyse des phénomènes, qui varient elles-mêmes selon les problématiques posées, les domaines et la finalité des études. Qu'il s'agisse de problématiques appliquées ou théoriques, l'objectif ultime est le même ; il s'agit de chercher à en établir une modélisation au sens propre du terme, c'est-à-dire un modèle simultanément définitoire, descriptif et prédictif du système complexe formé par l'ensemble des phénomènes collocationnels.

3 Les problématiques en linguistique appliquée

3.1 La didactique des langues

Nous regroupons sous l'appellatif « didactique des langues » à la fois l'apprentissage et l'enseignement des langues maternelles et secondes.

Concernant l'apprentissage des collocations en langue maternelle, nous ne disposons à l'heure actuelle d'aucune étude systématique. Si, du point de vue compréhension-interprétation, ces phénomènes ne semblent pas poser de problème majeur selon leur caractère intuitif, en revanche, du point de vue production, les erreurs ou problèmes dans leur maniement sont courants. Par exemple, chez les jeunes enfants (4-5 ans), la tendance est à la normalisation sémantique cumulée à une certaine proximité phonétique des unités lexicales substituées.

Ex : *vaisseau spécial* pour *vaisseau spatial*

Ex : *pot d'échauffement* pour *pot d'échappement*

Ex : *faire un coin d'œil* pour *faire un clin d'œil*

Ex : *à la fin du pestacle, on s'allume* pour *à la fin du spectacle, on salue* (collocation textuelle paradigmatique, comme pour *médecin... hôpital*)

Pour les publics plus âgés, Grossmann et Tutin pensent que « la modélisation visée devrait donc à la fois préciser les types collocatifs à l'origine de difficultés en réception et en production et fournir des aides “en ligne” lors de la production d'écrits » [Grossmann et Tutin 2003, p. 10].

Concernant l'apprentissage des collocations en langue seconde, les études [Granger 1998 ; Howarth 1998] montrent qu'elles sont difficiles à acquérir et surtout difficiles à produire, même pour des apprenants de niveau avancé ayant été sensibilisés à ces phénomènes. « S. Granger plaide pour un enseignement contrastif, à partir d'une modélisation en langue maternelle permettant de prendre conscience du phénomène. » [Grossmann et Tutin 2003, p. 10]. Ce qui ne résout pas le problème de savoir comment les intégrer dans la pratique didactique. « Il s'agit là d'un domaine encore largement à défricher. La conception de ressources lexicales spécifiques pour les apprenants apparaît également indispensable et on remarquera que de nombreux projets lexicographiques dans le domaine des collocations ont été avant tout conçus à destination des apprenants en langue seconde. » [Grossmann et Tutin 2003, p. 10]. Par exemple, dans le domaine de l'enseignement et de l'apprentissage de langues intelligemment assisté par ordinateur (ELIAO), LogoTax développé par Ludewig, est une application qui essaie de franchir le fossé habituel entre dictionnaires classiques de mots et expressions usuelles dans un texte. Ce système s'adresse aux étudiants en allemand ayant acquis déjà un niveau avancé et les aide à composer leur propre dictionnaire collocatif en leur permettant de sélectionner et mémoriser des collocations qui les intéressent pour leur propre production de langage. En fouillant dans un corpus LogoTax essaie de trouver et de classer des références textuelles illustrant l'usage d'une collocation sélectionnée par l'étudiant et d'en générer ensuite une entrée [Ludewig 2001].

3.2 La lexicographie

Dans une perspective lexicographique, les lexicographes sont confrontés au problème du codage des collocations dans les dictionnaires. Ce codage est nécessaire compte tenu des difficultés que les collocations posent en didactique et en traductologie, par exemple, mais il est rendu difficile, premièrement à cause de la nature idiomatique et/ou idiosyncrasique (aspects syntaxiques et sémantiques) des phénomènes collocationnels et deuxièmement par le peu de modèles théoriques sur lesquels s'appuyer, capables d'en proposer une modélisation.

À la suite d'Hausmann [Hausmann 1989] et de Mel'cuk [Mel'cuk et al. 1984, 1988, 1992, 1999], la plupart des dictionnaires de collocations comme le *BBi Dictionary of English Word Combinations* [Benson et al. 1986] ou encore le *Dictionnaires des cooccurrences* [Beauchesne 2001] ont adopté un système de codage à partir de la base et non du collocatif. Par exemple, la collocation *peur bleue* est codée à partir de la base *peur*. Toutefois, l'accès à l'information par la base n'est pas une évidence. En espagnol, l'adjectif *solo* dans *café solo*, traduit en français par *café noir*, pourrait être décrit à l'aide d'une correspondance entre le sens *sans lait* et l'adjectif *solo* dans l'article lexicographique de *café*, sans qu'aucune information ne soit perdue [Alonso Ramos 2003, p. 45]. Mais, si Alonso Ramos souligne que les collocations ne peuvent être décrites uniquement dans l'article lexicographique de la base, elle précise également que l'article du collocatif ne saurait suffire, dans la mesure où le lexicographe n'a pas les moyens d'y décrire l'interaction des sens des deux constituants.

Les collocations ont donné lieu à beaucoup d'ouvrages ; à titre d'exemples, nous citons les plus connus : le *BBi Dictionary of English Word Combination* [Benson et al. 1986], la base de données de Fontenelle [Fontenelle 1997, 2003] et le *Dictionnaire explicatif et combinatoire* [Mel'cuk et al. 1984, 1988, 1992, 1999].

Le *BBi Dictionary of English Word Combination* est un dictionnaire de collocations destiné aux apprenants et aux usagers de l'anglais qui regroupe 18 000 entrées et 90 000 collocations. Préalablement à la consultation du dictionnaire, le Practical Guide to the BBI informe l'utilisateur que deux types de collocations sont recensés. D'une part, les collocations grammaticales, répertoriées sous leur mot fort, sont divisées en huit classes tenant compte de phénomènes comme la transitivité, la passivation... D'autre part, les collocations lexicales, divisées en sept classes, formées par des enchaînements du type verbe + nom, adjectif + nom, nom + verbe... Les auteurs tiennent compte des phénomènes de polysémie. Ils donnent également des conseils pour trouver les réponses de façon optimale. Par exemple, concernant une collocation lexicale, si elle est composée d'un nom, chercher en premier lieu ce nom, si elle est composée de deux noms, chercher en premier lieu le deuxième nom, en l'absence de nom, chercher l'adjectif...

Dans le domaine de la lexicographie computationnelle et plus précisément de la lexicographie informatisée, la base de données de Fontenelle [Fontenelle 1997, 2003] transforme le dictionnaire Collins-Robert en base de données enrichie sémantiquement, afin de rendre accessible l'ensemble des connaissances métalinguistiques et collocationnelles. Dans cet ouvrage, l'auteur rend explicite l'organisation structurelle d'un domaine sémantique en décrivant la façon dont les éléments sont liés et en se servant de la théorie Sens-Texte [Mel'cuk 1997]. En utilisant l'appareil descriptif des 56 fonctions lexicales standard, l'auteur a codé environ 70 000 paires d'items linguistiques, en montrant parfois les imprécisions de la théorie Sens-Texte et la nécessité de créer de nouvelles fonctions lexicales.

Le *Dictionnaire explicatif et combinatoire* (du Français contemporain) – DEC¹⁶ – est également un dictionnaire de collocations, mais élaboré dans une perspective lexicographique théorique, sous-tendu par le modèle des fonctions lexicales, lui-même inclus dans la théorie Sens-Texte. Selon cette théorie, un DEC doit fournir, de façon systématique et rigoureuse, les informations permettant à un locuteur de construire toutes les expressions linguistiques correctes de n'importe quelle pensée, et ce dans n'importe quel contexte : c'est un dictionnaire de production. En conséquence, les lexies françaises sont décrites sous toutes leurs facettes : sémantique, syntaxique, lexico-combinatoire, morphologique, etc., d'où une décomposition de chaque article en pas moins de dix zones.

Pour chaque lexie à décrire, il s'agit de donner la définition, d'indiquer les connotations, de spécifier la cooccurrence syntaxique (c'est-à-dire les données du régime syntaxique avec tous les détails des conditions d'utilisation contextuelles), de recenser la cooccurrence lexicale restreinte au moyen des fonctions lexicales, sans oublier les indications sur la flexion, les marques d'usage ainsi que les spécifications orthographiques, prosodiques, pragmatico-culturelles et même encyclopédiques.

Quatre volumes du DEC ont été publiés, ce qui correspond à la description formelle de 510 vocables du français. Chaque volume est divisé en deux parties, l'une théorique, l'autre descriptive. La partie théorique comporte des textes présentant les principes, règles et critères qui régissent les descriptions lexicales ou justifiant la démarche adoptée, ainsi que des analyses des phénomènes linguistiques pertinents. La partie descriptive constitue le dictionnaire à proprement parler : les descriptions lexicographiques formelles y sont présentées. Toutefois, le métalangage du DEC et sa formalisation extrême le rendent peu susceptible d'un emploi grand public. C'est pourquoi des versions simplifiées comme le *DAFLES* [Verlinde et al. 2003] et le dictionnaire de combinatoire – *DiCo*¹⁷ [Polguère 2000b, 2003, 2005] sont en cours d'élaboration. Le DiCo est une base de données lexicale du français, dont la finalité première est de décrire chaque lexie selon deux axes : les dérivations sémantiques qui la lient à d'autres lexies de la langue et les collocations qu'elle contrôle. Cette description s'accompagne d'une modélisation des structures syntaxiques régies par la lexie et d'une modélisation de son sens, sous forme d'étiquetage sémantique. Au-delà de la construction d'une base de données décrivant le lexique du français, le projet DiCo vise l'élaboration de deux types de produits lexicographiques. D'abord, la production automatique à partir du DiCo (par compilation de celui-ci) de lexiques de systèmes de TAL [Lareau 2002 ; Steinlin et al. 2004], ensuite la création de versions « vulgarisées » du DEC, à l'usage du grand public, objectif poursuivi par le projet intitulé lexique actif du français – LAF [Polguère 2000a].

¹⁶ Cette présentation du DEC reprend la présentation générale proposée sur le site de l'université de Montréal : <http://www.olst.umontreal.ca/decfr.html>

¹⁷ Le DiCo est consultable en ligne à l'adresse suivante : <http://olst.ling.umontreal.ca/dicouebe/>

3.3 La terminologie

Traditionnellement, partant des notions propres à un domaine (finances, bourse, informatique...), la tâche du terminologue est de dresser « des listes d'unités lexicales rassemblées sous un terme de nature nominale » et de les répartir « soit par catégories grammaticales, soit par sens » [L'homme 2003, p. 90], pour en dresser un répertoire. De ce fait, les travaux en terminologie accordent plus de place à la construction de taxinomies qu'à l'aspect syntagmatique des mots ; d'où un intérêt minime pour les phénomènes phraséologiques, parmi lesquels figurent les collocations. Toutefois, depuis quelques années, plusieurs recherches portent sur les phénomènes combinatoires (syntaxiques et sémantiques) dans les textes spécialisés [Mortchev-Bouveret 1998]. Deux grandes tendances se dégagent ; l'une qui vise à recenser et à classer les cooccurents dans un domaine particulier [Cohen 1986] ou à rendre compte du fonctionnement des termes en incluant dans les entrées terminographiques des phénomènes relevant de la combinatoire lexicale [Binon et al. 2000 ; Meynard 2000 ; Caigon 2001] ; l'autre qui cherche à caractériser ces groupes lexicaux dénommés combinaisons lexicales spécialisées – CLS – [Heid 1994 ; Laporte et L'homme 1997] et/ou à proposer des repères méthodologiques pour les situer par rapport aux collocations (terme réservé aux lexicographes) [L'homme 2000]. Comme pour les lexicographes, dans ce dernier cas se pose à nouveau le problème du codage ou de l'encodage des CLS sous la base ou le cooccurrent [L'homme 2003]. Cette problématique est d'autant plus présente qu'avec le développement de banques de données terminologiques informatisées [Termium ; *Grand dictionnaire terminologique* – GDT – ; *Eurodicautom*] les rubriques consacrées aux associations lexicales peuvent désormais rendre compte des multiples possibilités combinatoires de certains cooccurents, travail facilité du point de vue théorique par la modélisation possible des collocations.

4 Les problématiques en linguistique théorique

4.1 La modélisation

4.1.1 Pourquoi et comment modéliser les collocations ?

« La modélisation des collocations, même si elle se révèle délicate, est une étape obligée. » [Grossmann et Tutin 2003, p. 8]. De fait, elle est nécessaire d'un point de vue applicatif pour l'apprentissage des langues secondes, la lexicographie, la terminologie ou encore le TAL, nous venons de le voir. Elle est également indispensable d'un point de vue théorique pour mieux appréhender les phénomènes collocationnels sur le plan linguistique. Le caractère idiomatique et/ou idiosyncrasique des collocations posé par le problème de leur non-compositionnalité sémantique nécessite de mettre en évidence simultanément les caractéristiques et les mécanismes qui régulent les associations syntaxiques et sémantiques. D'une façon générale, une modélisation « peut se définir comme une description formelle, explicite et cohérente des propriétés syntaxiques et sémantiques des collocations » [Grossmann et Tutin 2003, p. 10].

En particulier, une modélisation devrait être adéquate à la fois au niveau descriptif et au niveau explicatif. Sur le plan descriptif¹⁸, il s'agit de disposer :

- d'une description sémantique fine de la collocation (définition) et de ses propriétés syntaxiques.

Ex : pour *peur bleue*, la définition doit être accompagnée de données syntaxiques mentionnant le caractère inséparable du nom et de l'adjectif ;

- d'une description de la relation sémantique liant le collocatif à la base, au moins dans les cas où la relation est partiellement compositionnelle.

Ex : pour *peur bleue*, la relation est intensive (mais de quel type ? Sur quel composant de la lexie repose-t-elle ?) ;

- d'une description de la relation syntaxique liant le collocatif à la base.

Ex : pour *peur bleue*, *bleue* est un modifieur dépendant de *peur*.

Sur le plan explicatif, les collocations étant difficilement prédictibles mais la plupart du temps facilement analysables, il s'agit de pouvoir rendre compte des procédés sémantiques d'association entre les deux composants.

4.1.2 Le modèle des fonctions lexicales – FL

Mel'cuk signale que, dans le modèle Sens-Texte, « la définition proposée pour la collocation couvre TOUS les cas d'expressions qui sont linguistiquement particulières et nécessitent l'inclusion dans le dictionnaire sous l'entrée de leur base, et SEULEMENT de tels cas : elle est donc nécessaire et suffisante » [Mel'cuk 2003, p. 26]. Il reprend la terminologie d'Hausmann concernant le constituant libre d'une collocation qu'il appelle également la *base* et l'autre constituant, le *collocatif* (sélectionné en fonction du premier). Dans ce modèle, les collocations sont des expressions semi-figées appelées semi-phrasèmes (*café noir*, *commettre un crime*) qui s'opposent à des locutions complètement ou partiellement figées appelées phrasèmes complets (*pas question*, *prendre une veste*) ou quasi-phrasèmes (*point-virgule*, *bande dessinée*) et aux pragmatèmes (*bonne nuit*, *à consommer avant...*). « La combinatoire lexicale restreinte constitue un élément fort problématique pour toute description dictionnaire et, par ricochet, pour la linguistique théorique. Le caractère capricieux et imprévisible de la cooccurrence lexicale est notoire. » [Mel'cuk 1997, p. 23]. Or l'approche Sens-Texte propose un modèle fonctionnel de ce phénomène, les fonctions lexicales – FL : « Puisque le trait définitoire central de collocation est la dépendance fonctionnelle du collocatif vis-à-vis de la base, il est évident que les fonctions lexicales [= FL] représentent LE moyen pour la description des collocations. » [Mel'cuk 2003, p. 27].

Une FL se définit comme « [...] une dépendance, ou correspondance, **f** qui associe à une unité lexicale **L**, appelée l'ARGUMENT de **f**, un ensemble d'unités lexicales **f(L)** – la VALEUR de **f**. Chaque FL correspond à un sens très général (à la limite qui peut être zéro) et à un rôle

¹⁸ Nous reprenons les trois critères proposés par Grossmann et Tutin [Grossmann et Tutin 2003, p. 11].

syntactique profond » [Mel'cuk 1988, p. 6]¹⁹. Les fonctions lexicales correspondent à des fonctions mathématiques dont la formule est du type : $f(x) = y$, où x est l'argument ou le mot clé et y , la valeur. Il existe trois binômes de FL, décrits en fonction de trois axes.

4.1.2.1 Fonctions lexicales paradigmatiques vs syntagmatiques

Les FL paradigmatiques vs syntagmatiques se distinguent selon le sens associé à f et le comportement syntaxique des éléments de $f(L)$. Les FL décrivent les propriétés combinatoires des unités lexicales sur les axes paradigmatiques et syntagmatiques (les cooccurrences lexicales restreintes, c'est-à-dire les collocations).

Une FL f paradigmatique s'utilise dans le texte – au lieu de – L .

Ex : *fumeur*, qui est $S1(FUMER)$, s'utilise dans le texte au lieu du verbe *fumer*, permettant la paraphrase il *fume* = il est un *fumeur*

Les FL paradigmatiques modélisent donc des dérivations sémantiques de L , au sens propre (nom d'agent, nom de propriété, adjectif relatif...), comme au sens figuré (synonymes, antonymes...), et à ce titre ne sont donc pas pertinentes pour la description des collocations.

Une FL f syntagmatique s'utilise dans le texte – à côté de – L .

Ex : $Magn(FUMER)$ = *comme un sapeur* s'utilise dans le texte à côté du verbe *fumer*, permettant l'énoncé il *fume comme un sapeur*

Les FL syntagmatiques modélisent donc la cooccurrence lexicale restreinte de L (intensifieur, verbes supports...) et, à ce titre, entrent directement dans la description des collocations.

4.1.2.2 Fonctions lexicales standard vs non standard

Les FL standard et non standard mobilisent un critère d'universalité de f .

Une FL standard est (en principe) universelle, au sens où elle apparaît dans toutes les langues. À ce titre, elle s'applique à beaucoup de mots clés, possède énormément d'éléments comme valeur de FL et participe au paraphrasage en tant qu'objet des règles universelles de paraphrase. Les FL standard ont des noms latins universels (*Magn*, du latin *magnus*, qui signifie *grand*, FL appelée *intensificateur*) et sont désignées par des symboles spéciaux. Il en existe une soixantaine (57 exactement).

Ex : $S1(L)$: nom d'agent (*fumeur* de *fumer*) ; $Magn(L)$: intensificateur = *fumeur invétéré*

Ex : $Oper1(L)$: verbe support (*porter plainte*)

En principe, une FL non standard n'apparaît pas (mais le peut !) dans toutes les langues. À ce titre, elle s'applique à peu de mots clés (voire à un seul), possède également peu d'éléments comme valeur de FL (voire un seul) et ne participe au paraphrasage – en ce sens que les règles universelles de paraphrase n'utilisent aucune FL non standard. Les FL non standard sont décrites par des minidéfinitions dans la langue où elles s'appliquent.

Ex : sans produit laitier(CAFE) = [café] *noir*

Ex : première chez un enfant qu'il perd vers l'âge de 7 ans(DENT) = [dent] *de lait*

Ex : qui tend à changer(TEMPS) = [temps] *incertain*

¹⁹ Pour une définition formalisée de fonction lexicale, consulter [Mel'cuk 2003, p. 28 ; Kahane et Polguère 2001].

4.1.2.3 Fonctions lexicales simples vs complexes vs configurations de FL

Les FL simples vs complexes mobilisent un critère relatif à la structure syntaxique de **f**.
 « Une FL simple est constituée d'un seul nom de FL, tandis qu'une FL complexe représente une série de FL liées syntaxiquement qui ont le même mot clé et sont exprimées comme un tout ; une configuration de FL est une série de FL qui ne sont pas syntaxiquement liées, mais qui ont le même mot clé et sont exprimées aussi comme un tout. » [Mel'cuk 2003, p. 29].

Ex : FL simples standard : Magn(malade) = très, gravement

Ex : FL complexes : CausFunc1(CONFUSION) [= (causer que Func1) : *semer la confusion*]

Ex : Configurations de FL : [Magn + IncepReal1](LARMES) [= (commencer à réaliser intensément) : *fondre en larmes*]

4.1.2.4 Limites du modèle des fonctions lexicales

Pour les FL syntagmatiques en particulier, le modèle des FL vise une description exhaustive, systématique et transparente des collocations contrôlées par chaque lexie vedette. En ce sens, le modèle des FL satisfait aux exigences imposées par le principe d'une modélisation : d'une part, les FL permettent de décrire formellement la relation syntaxique et sémantique qui lie le collocatif à la base, d'autre part, la description sémantique de la collocation est déduite de la description sémantique proposée par la FL. Toutefois, sans remettre l'apport fondamental pour la description des collocations que représentent les FL, Mel'cuk lui-même [Mel'cuk 2003] ne cache pas que le modèle comporte encore à l'heure actuelle quelques imperfections. Premièrement, si l'extrême majorité des collocations peut être décrite par les FL, il demeure un petit groupe de collocations qui sont couvertes par le régime du mot clé (de la base), et non par des FL de la lexie vedette.

Ex : *assurance vie*, où la *vie* est ce qui est *assuré*, donc un actant sémantique et syntaxique de ASSURANCE.

Deuxièmement, toutes les collocations décrites par des FL n'utilisent pas toujours des FL standard. Un ensemble important de collocations est couvert par les FL non standard, qui ne peuvent être théoriquement prévues et doivent donc être recueillies empiriquement, par une recherche minutieuse.

Au niveau de la description syntaxique, il arrive par exemple que la description proposée pour certains verbes supports ne soit pas suffisante. Comme leur nom l'indique, les verbes supports servent de support à des prédicats nominaux [Gross 1981, 1991, 1993, 1994]. Leur signification est générale ; ils n'ont pas de fonction prédicative. Ils apportent aux substantifs prédicatifs qu'ils accompagnent des informations relatives au temps, à la personne, au nombre, etc. Des verbes comme *être*, *avoir*, *faire* ou *donner* ont généralement un emploi de verbe support. Le verbe support permet d'actualiser le prédicat.

Ex : *être heureux* ; *avoir peur* ; *faire un voyage* ; *donner une gifle*

Par voie de conséquence, Gross parle également de noms supports [Gross 1986].

Ex : *douanier* → *agent des douanes*

Dans ces cas, le modèle des FL rencontre des difficultés à indiquer certaines informations relatives à la passivation, à la modification du nom ou encore aux changements de déterminants. Pourtant, « le traitement de ces phénomènes serait grandement amélioré par

un traitement plus systématique des propriétés syntaxiques à partir du collocatif » [Grossmann et Tutin 2003, p. 13 ; Tutin 1997].

Enfin, la littérature fait état de plusieurs remarques allant du métalangage, parfois complexe et opaque [Wanner 1996], aux règles de combinaisons et de configurations fantaisistes en passant par l'amélioration du marquage des registres et des niveaux de langue et de familiarité des lexies.

Disposant d'un modèle de formalisation possible des collocations, une autre problématique a trait à la motivation des deux éléments. En effet, au niveau sémantique, un intérêt doit être porté à l'étude des mécanismes permettant de générer les collocations.

4.2 La motivation sémantique

La question de la motivation sémantique des collocations est souvent évacuée dans les modèles lexicaux, tout en étant pourtant centrale en linguistique. Dans de nombreuses collocations, l'arbitraire n'est pas total, et le caractère analysable de l'association entre la base et le collocatif peut être catégorisé en deux grands types de motivations : les procédés sémantiques productifs et les régularités sémantiques.

4.2.1 Les procédés sémantiques productifs

Les procédés sémantiques productifs correspondent aux cas de la métaphore, de la comparaison et de la métonymie et concernent de ce fait les collocations imagées.

Ex : *rouge de colère ; faim de loup ; océan de bonheur*

Ces motivations « ont leur source dans le principe cognitif permettant d'appréhender un phénomène sous l'angle d'un autre, dans l'ancrage anthropologique de la représentation, ou encore dans des stéréotypes sociaux » [Grossmann et Tutin 2003, p. 14]. Par exemple, nombreuses sont les collocations construites par métaphores empruntées au domaine du bâtiment pour caractériser l'argumentation [Lakoff et Johnson 1985].

Ex : *construire une argumentation ; les fondements de son argumentation*

Ces procédés stylistiques sont en effet largement utilisés dans les associations collocationnelles. De fait, l'expression de la collection et de la quantification, l'appréciation méliorative ou dépréciative, l'intensité [Tutin et Grossmann 2002]... sont générées grâce à des moules productifs qui reposent sur certains types de métaphores ou de comparaisons. Les collocations créées de cette manière sont donc partiellement régulières, et il est possible de préciser quelques-unes de leurs règles de formation.

En français, la collection et la quantification sont exprimées par un syntagme binominal en *de* dont le collocatif, le premier nom est collectif et revêt une valeur quantitative importante, d'où la valeur métaphorique. Dans ce cas, la valeur métaphorique tient au fait que le nom collectif perd son autonomie référentielle.

Ex : *une forêt de mâts ; une armée de marmitons ; un essaim d'hélicoptères*

Le choix du nom collectif n'est pas arbitraire et suit certains critères d'appropriation référentielle (**essaim de mâts*).

Selon le même principe, il existe aussi des noms qui n'ont pas de valeur collective mais qui ont également une valeur quantitative acquise métaphoriquement.

Ex : *une avalanche de questions ; un déluge de paroles*

Pour pouvoir prédire ces associations, il faut préciser les conditions d'application référentielle liées au sémantisme des expressions. Un quantitatif d'origine métaphorique « semble pouvoir être considéré comme un outil permettant d'exprimer la quantité, chaque fois qu'il est question d'« objets » (abstraits ou concrets) qui nous « tombent dessus » en grande quantité, soudainement et par surprise » [Tutin et Grossmann 2002, p. 16]. Les collocatifs quantitatifs sont donc sélectionnés en fonction de la cohérence de l'élément métaphorique avec le signifié de la base.

De même que pour la quantification, l'appréciation méliorative ou dépréciative se construit avec un syntagme binominal en *de*. Toutefois, à la différence du premier cas, cette fois, le collocatif est en deuxième position. Il suit la préposition pour marquer l'appréciation positive ou négative, comme dans *une santé de fer, un morceau de roi, un caractère de cochon*. « Les appréciatifs, fortement liés à des stéréotypes culturels ou à des informations encyclopédiques (*une voix de stentor, une voix de sirène*), ont un comportement différent, *a priori* moins facilement prédictible. » [Tutin et Grossmann 2002, p. 17]. Lorsque le collocatif introduit est un nom propre, le mécanisme stylistique mobilisé correspond généralement à une comparaison, qu'il s'agisse de structure binominale ou construite avec des adjectifs relationnels.

Ex : *la force d'Hercule ; une force herculéenne*

On trouve également des structures en *comme* pour évoquer l'appréciation.

Ex : *fort comme un turc*

Quelle que soit la structure syntaxique retenue, on se trouve soit dans le cas de collocations transparentes, lorsque la signification de la comparaison est largement partagée et repose sur des stéréotypes culturels que la langue a lexicalisés dans de nombreuses expressions (fer, acier = force), soit dans le cas de collocations opaques, lorsque la motivation est de type historique ou mythologique, et la collocation est forcément plus idiosyncrasique. Dans ce dernier cas, c'est plutôt l'adjectif qui sélectionne sémantiquement le nom.

Ex : *de bénédictin* sélectionne de manière standard *travail* et autorise *minutie*

Dans ce cas, il s'agit de collocations régulières, la signification de *minutie* étant incluse dans celle de *bénédictin*.

Concernant l'intensité, le système de production des collocatifs intensifs est très proche des appréciatifs. On retrouve, entre autres, des adjectifs relationnels, des structures binominales en *de* et des comparaisons en *comme*.

Autant pour les adjectifs relationnels la prédictibilité est possible et apparaît relativement transparente (**Ex :** *un froid sibérien ; un froid polaire*), autant pour les structures binominales et les comparaisons la situation est plus complexe. Tantôt la motivation est accessible et donc prédictible. Par exemple dans les structures binominales où le collocatif correspond à un nom d'animal incarnant un défaut (**Ex :** *têtu comme une bourrique*) ou une qualité (**Ex :** *malin comme un singe*) ; dans ce cas, la motivation est sous-tendue par un symbolisme culturel. De même, les cas de comparaisons en *comme*, où le collocatif est un intensifieur.

Ex : *blanc comme neige*

À l'inverse, il est des collocations fondées sur des structures binominales imagées qui sont imprédictibles. Par exemple, le collocatif peut être un nom d'animal, sans que celui-ci n'incarne de qualité « culturellement fondée » (**Ex** : *une mémoire d'éléphant* ; *un froid de canard*). De même, il est des cas de comparaisons quasiment figées qui sont idiomatiques (**Ex** : *un temps à ne pas mettre un chien dehors*) et donc imprédictibles.

Le problème est que la cohérence de tels réseaux est difficile à établir, ce qui signifie que ces collocations sont peu prédictibles. Premièrement, en général, plusieurs réseaux sémantiques sont simultanément mobilisés. Lakoff et Johnson montrent que l'argumentation est décrite par des métaphores liées à l'opposition fond vs surface, mais aussi par des métaphores de forme comme dans *les grandes lignes de l'argumentation*. Deuxièmement, la motivation est souvent obscurcie par des phénomènes liés à l'usage. Certaines collocations métaphoriques ou métonymiques perdent la transparence de leur motivation par l'usure et leur appartenance à un style noble « dépassé », créant un décalage de désuétude.

Ex : *vie orageuse* ; *remords dévorant*

De même, les collocations à valeur intensive construites par comparaisons ont tendance à être familières, sinon vulgaires.

Ex : *mauvais comme la gale* ; *con comme la lune*

Dans ces cas, le collocatif est sémantiquement motivé mais ne peut être intégré dans un réseau, car il est idiosyncrasique.

Face à ces limites, « un programme cohérent en la matière viserait à mettre en évidence, pour les collocations imagées, les réseaux collocationnels les plus productifs, à examiner la diversité de leurs réalisations syntaxiques et à spécifier les restrictions d'utilisation » [Grossmann et Tutin 2003, p. 15].

Ainsi, beaucoup de collocations sont construites par processus figuraux comme la métaphore, la métonymie ou encore la comparaison. Il apparaît que, pour les collocatifs à valeur quantitative, la motivation sémantique est forte, tandis que, pour les collocatifs à valeur appréciative ou intensive, bien qu'ils recourent à des structures régulières, le caractère prédictif du collocatif n'est pas toujours assuré [Tutin et Grossmann 2002, p. 17].

4.2.2 Les régularités sémantiques

« Bien qu'on ait souvent insisté sur le caractère irrégulier des collocations, il apparaît possible, dans de nombreux cas, de préciser les règles d'association entre la base et le collocatif, en fonction de leur sémantisme respectif. » [Grossmann et Tutin 2003, p. 15]. En particulier pour les collocations les moins imagées, les régularités sémantiques correspondent à des traits sémantiques récurrents [Grossmann et Tutin 2005].

Ex : *affreux* sens intensif avec des noms de sentiments à polarité négative comme *désespoir*

Dans leur étude sur les intensifieurs des « noms de joie », Grossmann et Tutin [Grossmann and Tutin à paraître] dégagent trois principes explicatifs des associations entre adjectif intensifieur – AdjIntens – et nom de joie – Ndj – : le renforcement, la convergence de polarité et la cohérence avec le type sémantique du Ndj. Ces trois types de motivations résident dans l'idée d'une « congruence sémantique », pour reprendre le terme des auteurs, ce qui correspond à une logique des signifiés lexicaux qui tendanciellement favorise les affinités. La congruence sémantique requiert que les dimensions sémantiques repérées pour le nom et pour

l'adjectif ne soient pas contradictoires. Dans la mesure où l'idée de *bonheur* semble, de façon privilégiée, être liée à la durée, les collocatifs seront plutôt du type *paisible, calme*, s'opposant à l'idée même d'intensité, telle qu'elle peut s'exprimer à travers des adjectifs comme *vif, aigu* ; cela n'empêche évidemment pas des combinaisons du type *vif sentiment de bonheur* dans lesquelles *vif* caractérise *sentiment* et non *bonheur*. Par ailleurs, même pour les collocations les moins imagées, on retrouve parfois les réseaux métaphoriques. Par exemple la métaphore du fond vs surface, structure de nombreuses expressions.

Ex : *la profondeur des sentiments ; être profondément ému*

Le renforcement concerne par exemple les cas où le nom comporte déjà le trait intensif et où l'utilisation de l'adjectif intensif conduit à un renforcement, par l'utilisation d'un intensif adjectival de degré élevé comme un superlatif.

Ex : *tristesse* est associé à *grande* ; *désespoir* est associé à *immense* ou encore à *total* et non à *grand*, parce que le *désespoir* est *grand* par définition, donc l'intensif doit être plus fort qu'avec *tristesse*, par exemple.

Les auteurs illustrent la convergence de polarité avec le trait *agréable*. Si la base comporte le trait *agréable*, le collocatif revêt en principe le trait appréciatif (au sens mélioratif) ou un trait neutre. À l'inverse, si la base comporte le trait de polarité désagréable, le collocatif revêt le trait dépréciatif ou neutre. Par exemple, *affreux* est un adjectif de polarité dépréciative. Une étude sur corpus a montré qu'il n'exprime l'intensité qu'en cooccurrence avec un substantif orienté négativement.

Ex : *affreux désespoir* ; **affreux bonheur* les auteurs précisent que cette combinaison est possible, stylistiquement marquée, mais que dans ce cas l'adjectif perd son statut d'intensif.

Là encore, une étude sur corpus a également montré que dans certains cas, en particulier lorsque l'adjectif est antéposé, le trait de polarité positive s'efface.

Ex : *parfait bonheur* ; *parfait désespoir* dans ces deux cas, les auteurs précisent que le sens de *parfait* qui est mobilisé correspond à la signification *total* et non à la signification *excellent*.

Enfin, le principe de cohérence avec le type sémantique du Ndj est un principe plus flou. Il concerne les cas où une base conserve une ambivalence d'ordre sémantique telle que la possibilité de dénoter une émotion, un état ou une qualité, par exemple. Dans ce cas, ce principe d'association montre que « certaines contraintes de sélection liées aux AdjIntens peuvent favoriser ou bloquer l'association avec ces noms polysémiques » [Grossmann and Tutin à paraître].

Ex : *absolu* ou *total* s'associent préférentiellement à un Ndj traduisant un état comme *bonheur* ou *désespoir* mais semblent incompatibles avec un Ndj comme *chagrin*.

La question de la motivation sémantique des collocations est donc traitée selon deux aspects, les procédés sémantiques productifs et les régularités sémantiques. Les trois mécanismes décrits dans les régularités sémantiques, notamment le renforcement et la convergence de polarité, sont également pertinents dans la perspective d'une description argumentative de la motivation des entités mobilisées dans certains phénomènes collocationnels.

5 Une motivation argumentative

5.1 Problématique et filiation de la recherche

Adhérant au point de vue de Tutin et Grossmann concernant la nécessité d'étudier les mécanismes permettant de générer les collocations. L'idée de décrire argumentativement la motivation des entités mobilisées dans certains phénomènes collocationnels provient du pouvoir explicatif et prédictible²⁰ « potentiel » de la sémantique des possibles argumentatifs, modèle lexical de description de la signification (*cf.* Chapitre 2 sur la sémantique des possibles argumentatifs). La signification lexicale y est conçue comme une entité pluridimensionnelle : sa description mobilise simultanément des informations sémantiques, internes à la langue, et des informations discursives, externes donc contextuelles. Le caractère mixte du modèle repose sur une interaction nécessaire entre le sémantique et le discursif. Aussi, le recours à la sémantique des possibles argumentatifs inscrit logiquement cette recherche dans la *filiation contextualiste*. En effet, de notre point de vue, l'approche contextualiste des phénomènes collocationnels revêt également un caractère mixte, puisque l'analyse des collocations se fonde sur l'observation empirique des relations sémantiques – dimension sémantique –, créées par le contexte – dimension discursive –. Rappelons que lorsque Firth [Firth 1957] introduit pour la première fois le terme collocation, il entend caractériser certains phénomènes linguistiques de cooccurrence qui engendrent des modifications de sens constatées sur les mots par leur relation de collocation en contexte. L'approche contextualiste étudie les collocations dites textuelles, eu égard au phénomène de cohésion textuelle, inhérent à toute unité sémantique formée par le texte, dont ces collocations rendent compte et auquel elles participent. Pour les mêmes raisons, l'hypothèse d'une motivation argumentative des entités collocationnelles est propre à désigner les collocations textuelles comme objet d'étude.

5.2 Définition retenue

5.2.1 Les collocations textuelles

Dans la perspective contextualiste et néo-contextualiste, la définition de collocation textuelle mobilise le sens premier du concept de collocation, c'est-à-dire qu'elle désigne des relations de sens portées par des mots qui tendent à apparaître ensemble dans un contexte textuel donné. Firth et Sinclair considèrent que seules les collocations lexicales méritent l'appellation de collocation. Pour notre part, les colligations étant constituées de mots grammaticaux dénués de contenu notionnel, elles sont impropres à une description argumentative fondée sur la signification des entités qui la composent. Aussi, nous laissons les collocations grammaticales de côté, au profit des collocations lexicales, construites avec des mots pleins, susceptibles de contenir des enchaînements argumentatifs séquentiels. Par ailleurs, les relations analysées étant d'ordre argumentatif, *a fortiori* pragmatique, nous considérons également que le critère de bonne formation syntaxique n'a pas lieu d'être.

²⁰ *Prédictible* est employé au sens de *calculable* et rejoint par ce fait, l'une des perspectives de recherches mentionnées par Mel'cuk [Mel'cuk 1997, p. 23], le *facilement* en moins (*cf.* Présentation de ce chapitre) !

Les développements successifs de la linguistique de corpus ont eu pour effet d'éloigner progressivement les critères liés à la compétence des locuteurs natifs, aux aspects sémantiques et syntaxiques, tout en privilégiant l'instanciation statistique de la collocation textuelle. Dans la lignée de Kilgarriff, Sinclair, Hoey (etc.), nous attribuons au critère statistique un caractère discriminant, sans toutefois recourir aux mesures statistiques t-score et IM, généralement réservées aux procédures d'extraction automatique, ce qui n'est pas le cas dans ce travail. Au contraire, le critère statistique a fait lui-même l'objet d'une révision dans une perspective qualitative (*cf.* Chapitre 4 sur le protocole de validation). Conformément aux analyses statistiques menées par Jones et Sinclair, les collocations jugées significatives sont retenues dans un espace de texte court qui correspond à une fenêtre de plus ou moins 4 mots autour du nœud. Le terme « nœud » n'est employé que pour désigner l'unité recherchée par concordance. Contrairement à la terminologie contextualiste généralement utilisée, nous conservons les termes de base et de collocatif introduits par Hausmann, précisément parce que le rôle sémantique joué par la base intervient dans les processus sémantico-discursifs analysés.

À cette acception statistique de la collocation textuelle, nous ajoutons une distinction fondée sur une typologie sémantique.

5.2.2 Les collocations libres *vs* restreintes

Les typologies sémantiques dédiées à la description – classification des collocations textuelles sont généralement larges, au sens où les distinctions qu'elles impliquent couvrent des phénomènes collocationnels très différents. Nous reprenons celle de Cowie (*cf.* Section 2.3.2.3.3), qui dissocie les collocations libres des collocations restreintes. La première catégorie inclut des associations de mots dont les deux éléments conservent leur sens littéral et dont la combinatoire est assez libre pour sembler quasi illimitée. La seconde catégorie inclut, de façon traditionnelle, des associations de mots dont l'un des constituants conserve son sens littéral tandis que l'autre a un sens figuré ; dans ce cas, le sens du mot A limite les possibilités combinatoires de l'élément B.

Les problématiques de recherches en linguistique théorique montrent que dans de nombreuses collocations, l'arbitraire n'est pas total. Les collocations libres sont même définies par le caractère prédictible de l'association entre la base et le collocatifs : « ...Cooccurrence that are free and predictable in the sense that they can be formed without any help by learner of German who knows these lexical items and who has mastered the elements of German grammar... » [Hausmann 1979, cité dans Benson 1989, pp. 2-3]. Toutefois, cette prédictibilité n'a pas de réalité autre que « psychologique » ou « associative », au sens défini par Partington [Partington 1998]. Concernant les collocations restreintes, englobant les collocations opaques, transparentes et régulières définies par Grossmann et Tutin (*cf.* Section 2.3.1.2.1), les auteurs ont montré le caractère analysable de l'association entre la base et le collocatif par motivation sémantique. En particulier pour les collocations régulières, les processus sémantiques en œuvre correspondent à des traits sémantiques récurrents permettant également une forme de prédictibilité.

La sémantique des possibles argumentatifs permettant de formuler une hypothèse sur la nature *sémantico-discursive* de la relation qui unit la base au collocatif dans les collocations textuelles (*cf.* Chapitre 4 sur les hypothèses de recherche...), nous entrevoyons la possibilité de mettre à jour un même principe descriptif – explicatif, voire prédictible pour les collocations libres et les collocations restreintes. Ce faisant, les collocations libres ne seraient pas « réellement » libres, au sens où leurs associations seraient contraintes par un principe argumentatif, principe également sous-jacent à leur prédictibilité, autre que psychologique ; tandis que la prédictibilité des collocations restreintes reposerait sur un nouveau mécanisme de génération, autre que syntaxique et/ou sémantique.

Conclusion

À l'interface entre la langue et le discours, les collocations posent la question du statut de l'irrégularité dans le lexique. Leur caractère idiomatique et idiosyncrasique est à l'origine de la difficulté à les définir, difficulté accrue par la nature dynamique et le caractère fluctuant des phénomènes qu'englobe le concept de collocation. Aussi, les solutions de descriptions adoptées prennent la forme d'une série de critères discutés et discutables selon l'approche choisie, soit lexicologique – lexicographique, soit linguistique de corpus. Ces deux approches retiennent une acception différente du concept de collocation. La première est caractérisée de « restreinte », fondée sur des paramètres syntaxico-sémantiques, orientés vers une définition formelle du concept. La seconde est qualifiée de « large », reposant essentiellement sur des critères statistiques hérités d'une conception contextualiste du sens, à laquelle les collocations participent par la fonction cohésive qu'elles assurent au sein des textes. Chacune de ces approches privilégie le recours à ses propres typologies, compléments nécessaires à la délimitation et à la catégorisation des phénomènes observés. Outre les différentes approches, les variations de définitions et la diversité des critères permettant de les décrire, les collocations sont issues des différentes méthodes de traitement et d'analyse des phénomènes, qui varient elles-mêmes selon les problématiques posées, les domaines et la finalité des études. Qu'il s'agisse de problématiques appliquées au domaine didactique, lexicographique, terminologique, ou de problématiques théoriques, les objectifs convergent : il s'agit d'une part de chercher à établir une modélisation de l'ensemble des phénomènes collocationnels, c'est-à-dire de mettre au point un modèle simultanément définitoire, descriptif et prédictif, et d'autre part, de mettre à jour les mécanismes inhérents à la formation des collocations. De fait, la question de la motivation notamment sémantique des collocations est souvent évacuée dans les modèles lexicaux, tout en étant pourtant centrale en linguistique. Dans cette perspective, ce travail présente une proposition de description de l'association collocationnelle en termes de motivation argumentative. Cette proposition repose sur le pouvoir explicatif et prédictible « potentiel » de la sémantique des possibles argumentatifs, modèle lexical de description de la signification. Le caractère mixte du modèle reposant sur une interaction nécessaire entre le sémantique et le discursif, le recours à la sémantique des possibles argumentatif inscrit logiquement cette recherche dans la filiation contextualiste, et par voie de conséquence, l'hypothèse d'une motivation argumentative des entités collocationnelles est propre à désigner les collocations textuelles comme objet d'étude.

Chapitre 2

La sémantique des possibles argumentatifs

Présentation

La sémantique et la pragmatique sont deux disciplines ayant respectivement pour objet d'étude la signification lexicale et le sens, selon que l'on se situe dans la langue ou dans le discours. Historiquement, la triade scolastique – res-conceptus-vox –, reprise par Morris sous la forme – syntaxe-sémantique-pragmatique – [Morris 1938, 1964, 1971] limite la description de la signification lexicale à la sémantique : « On suppose en effet que chaque mot exprime, indépendamment des autres, une certaine notion, et que la syntaxe établit des rapports sémantiques entre ces notions, rapports qui constituent des « opinions », des objets de croyance. Que certaines de ces opinions soient socialement admises, et d'autres socialement rejetées, la linguistique n'en a que faire ; elle doit se contenter, ce n'est déjà pas si simple, de calculer en quoi consistent les unes et les autres. » [Carel et Ducrot 1999, p. 6]. Une telle conception de la sémantique linguistique limite le champ d'investigation du domaine, en l'empêchant de rendre compte de certains phénomènes comme les « expressions socialement paradoxales », pour reprendre les termes de Carel et Ducrot, expressions qui expriment une opinion contraire à l'opinion commune. Par répercussion, l'analyse de telles expressions est souvent relayée aux domaines de la pragmatique et de l'analyse du discours. Or les développements récents en sémantique argumentative montrent que, pour une description exhaustive de la signification lexicale, entendue comme pluridimensionnelle, l'interaction entre sémantique et pragmatique, entre langue et discours, semble possible sinon nécessaire. Dans cette perspective, ce chapitre, sans chercher à effectuer un état de l'art sur les modèles de description de la signification, du sens et de leur articulation, a pour vocation de présenter exhaustivement le modèle de la sémantique des possibles argumentatifs. Ce modèle mixte de description-construction-représentation-interprétation de la signification lexicale, à partir d'informations linguistiques et discursives, s'inscrit dans la filiation directe de la sémantique argumentative, à laquelle il emprunte préceptes et notions ; une présentation préalable de la sémantique argumentative est donc nécessaire.

1 La sémantique argumentative

1.1 La théorie de la polyphonie

La théorie de l'argumentation dans la langue, développée par Anscombre et Ducrot depuis 1973 [Ducrot 1972, 1973 ; Anscombre et Ducrot 1983], s'inscrit dans une théorie plus générale dite de l'énonciation. L'énonciation est entendue comme « [l'] activité langagière exercée par celui qui parle au moment où il parle. Elle est donc par essence historique, événementielle, et comme telle ne se reproduit jamais deux fois identique à elle-même » [Anscombre et Ducrot 1983, p. 36-37]. La description de l'énonciation permet l'accès au sens des énoncés. Les éléments nécessaires à cette description se répartissent en deux grandes catégories d'indications, les unes argumentatives, les autres illocutoires. Bien que nécessaires, ces indications sont secondaires comparées à des indications primaires portées par le sens même des énoncés, qui concernent le ou les auteurs éventuels de l'énonciation. La notion centrale de cette théorie est la *polyphonie*, concept *esquissé* par Ducrot dans *Le dire et le dit* [Ducrot 1984], que nous résumerons ainsi : lorsqu'un locuteur L produit un énoncé E, dans le même temps il met en scène un ou plusieurs énonciateurs accomplissant des actes illocutoires [Ducrot 1984, p. 171-233]. Dans la mesure où l'énonciateur se distingue du ou des locuteurs, il peut choisir de s'identifier à lui/eux (il prend en charge ses/leurs actes illocutoires) ou bien de s'en distancier (il rejette alors ses/leurs actes illocutoires sur quelqu'un d'autre, distinct de lui, qui peut être déterminé ou non). Il s'agit là de deux attitudes énonciatives possibles. Ces postulats engendrent un certain nombre d'implications théoriques concernant la réalisation des actes de langage et notamment à propos de l'identité de celui qui endosse la ou les responsabilités illocutoires associées à ces actes discursifs *via* l'énonciation. Les indications polyphoniques sur l'énonciation sont qualifiées de *primitives* par Ducrot [Ducrot 1984, p. 193], parce qu'elles interviennent préalablement aux descriptions argumentatives et illocutoires de l'énonciation, sans toutefois être nécessaires à l'élaboration de celles-ci. Aussi, nous ne développons pas davantage les fondements de la théorie polyphonique de l'énonciation, dans la mesure où les notions qu'elle mobilise ne viennent pas éclairer la théorie de l'ADL utile à notre recherche.

1.2 L'argumentation dans la langue

La théorie de l'argumentation dans la langue – ADL – est un modèle centré sur la signification qui se propose de décrire une phrase par les enchaînements argumentatifs possibles dans le discours à partir d'elle. Pour Anscombre et Ducrot [Anscombre et Ducrot 1983], décrire une phrase revient à lui attribuer une valeur sémantique. Selon eux, cette attribution relève non pas de l'observation mais de l'explication, « il s'agit d'attribuer à chaque phrase une signification telle que l'on puisse, à partir de cette signification, prévoir le sens qu'aura son énoncé dans telle ou telle situation d'emploi » [Ducrot 1980, p. 8].

Avant de présenter la progression théorique de l'ADL, nous rappelons ci-dessous, en les reformulant *a posteriori*, les définitions que proposent Anscombre et Ducrot [Anscombre 1989, 1995c ; Anscombre et Ducrot 1983 ; Ducrot 1972, 1980, 1984, 1995b] pour « langue », « phrase », « signification », « parole », « énoncé » et « sens ».

La langue correspond à la signification des entités linguistiques (mots, syntagmes, phrases).

La phrase est une entité de la langue ; elle est abstraite, purement théorique et regroupe un ensemble de mots combinés selon les règles de la syntaxe. Cet ensemble est nécessairement pris en dehors de toute situation de discours.

La signification d'une phrase est sa valeur sémantique ; elle peut rendre compte et/ou prévoir le sens qu'aura l'énoncé de cette phrase.

La parole – le discours – correspond au sens des énoncés, résultat d'actes singuliers dans des situations énonciatives inédites.

L'énoncé, composant de la parole, est opposé à la phrase ; c'est ce que produit un locuteur, ce qu'entend un auditeur, c'est-à-dire l'occurrence discursive particulière d'une phrase.

Le sens d'un énoncé correspond à la description que donne l'énoncé de son énonciation, c'est-à-dire de ses possibilités d'emploi. Le sens peut encore se distinguer de sa « signifiante », c'est-à-dire du sens qu'un individu donne à un énoncé par sa propre vision du monde.

Les préoccupations premières inhérentes au développement de l'ADL concernent la description « argumentative » de certains « mots du discours », à commencer par les connecteurs pragmatiques (**Ex :** *mais, donc, puisque, car, d'ailleurs, ne...que...*). Ces entités sont présentées comme ayant la particularité d'être dénuées de signification lexicale, privées de tout trait notionnel¹. En revanche, elles ont un sens instructionnel ; elles jouent un rôle discursif essentiel, servant à marquer le rapport du locuteur à la situation d'énonciation. Une description argumentative de ces entités consiste à établir la liste des faisceaux d'enchaînements argumentatifs qu'elles autorisent.

Ex : *Il pleut. Décidément, je n'ai pas de chance.*

Ducrot [Ducrot 1980, pp. 131-159] explique que *décidément* est indicatif d'une classe de situations qui implique l'existence d'une série de faits. Ces faits peuvent être linguistiquement contraires à l'attente de l'énonciateur. En tout cas, le dernier de ces faits déclenche, en vertu d'une cause commune et indéterminée, la conclusion sur la nécessité pour le locuteur d'énoncer *décidément* (*P*), où (*P*) est un énoncé que *décidément* accompagne².

En fait, si deux segments peuvent être liés dans le discours par l'un de ces connecteurs pragmatiques, ce n'est pas tant par le contenu référentiel que ces segments arborent que par ce que leur structure linguistique impose en termes de contraintes sur leur orientation argumentative.

Ex : *Il a un peu travaillé. Il va donc réussir.*

Ex : **Il a un peu travaillé. Il va donc échouer.*

Ex : *Il a peu travaillé. Il va donc échouer.*

Ex : **Il a peu travaillé. Il va donc réussir.*

Ces exemples fétiches montrent que les mêmes indications factuelles – le fait de travailler – peuvent être orientées vers des conclusions opposées selon leur « habillage linguistique ». En l'occurrence, les opérateurs argumentatifs *un peu* et *peu* modifient la visée argumentative de la phrase à laquelle ils sont respectivement appliqués. En introduisant l'un ou l'autre des opérateurs, une seule conclusion devient (théoriquement) possible. C'est pourquoi « les possibilités de coordinations argumentatives à partir d'une phrase se fondent, directement, sur la signification de cette phrase, sans passer par les faits auxquels la phrase peut référer.

¹ Ce point de vue est à relativiser dans la mesure où chacun de ces mots est défini dans n'importe quel dictionnaire de langue.

² Pour une description exhaustive de *décidément*, consulter [Ducrot 1980, Chap. 4].

Ce qui résume le slogan “ L’argumentation est dans la langue ” » [Ducrot et Schaeffer 1995, p. 563]. Si l’argumentation est bien dans la langue, les enchaînements argumentatifs possibles et les conclusions envisageables à la suite d’une phrase appartiennent, eux, à l’espace du discours.

Au fur et à mesure de leurs recherches, Anscombe et Ducrot radicalisent leur propos en considérant les phrases uniquement comme des instruments pour la construction du discours, négligeant ainsi totalement la dimension référentielle de celles-ci. Cette radicalisation théorique les amène à étendre cette description argumentative à tous les mots du lexique, c’est-à-dire y compris les mots lexicaux ayant un contenu notionnel. Ces mots seraient caractérisés non pas par leur signification intrinsèque mais par les enchaînements possibles concernant une phrase les contenant : « Le mot, conçu comme entité linguistique abstraite, ne collabore au sens de l’énoncé que d’une façon indirecte : il commence par se combiner aux autres pour constituer la signification de la phrase, et c’est celle-ci qui, vu la situation de discours, produit le sens de l’énoncé. » [Ducrot 1980, p. 9].

Ex : *Il travaille beaucoup. Il va donc réussir.*

Ex : *Il travaille beaucoup. Il est donc fatigué.*

Dans ces exemples, les enchaînements argumentatifs *réussir* et *être fatigué* sont possibles par la combinaison des mots *travailler* et *beaucoup*.

Toutefois, le paradigme contenant *un peu* et *peu* montre que décrire l’opposition entre ces deux opérateurs par des potentiels argumentatifs opposés ne fonctionne que partiellement.

Ex : *Il a un peu travaillé. Il va donc réussir.*

Ex : [L’examen est très difficile] **Il a un peu travaillé. Il va donc réussir.*

Ex : [L’examen est très difficile] **Il a peu travaillé. Il va donc réussir.*

Ex : **Il a peu travaillé. Il va donc réussir.*

Dans ce paradigme, l’opérateur *un peu* ne modifie pas la visée argumentative de l’énoncé auquel il est appliqué, les deux conclusions sont dites dans un premier temps asémantiques. Puis, dans un deuxième temps, le discours *Il a peu travaillé. Il va donc réussir* est jugé tout aussi dicible que le premier *Il a un peu travaillé. Il va donc réussir* et son caractère paradoxal ne réfère pas à une « impossibilité linguistique » [Carel et Ducrot 1999, p. 7]. C’est pour pallier ce type d’objection qu’ont été construites les notions de « topos » et de « forme topique ». Au lieu de soutenir que seul l’un des deux enchaînements est possible, Anscombe et Ducrot ont été amenés à dire que ces deux argumentations se fondent sur des principes argumentatifs – ou *topoi* – opposés. Le premier topos relie *travail* et *réussite*, le second, *travail* et *échec*. L’effet opposé de *un peu* et *peu* ne tient donc pas à ce qu’ils n’autorisent pas les mêmes conclusions mais à ce qu’ils ne permettent pas l’emploi des mêmes *topoi*. La mise en évidence de *topoi* dans les enchaînements argumentatifs conforte l’idée centrale de l’ADL, qui pose que les mots et les structures phrastiques contraignent les enchaînements argumentatifs indépendamment des contenus véhiculés par les énoncés en question : « Toutes les relations argumentatives entre énoncés [...] ne sont pas déductibles de leur contenu informatif. » [Anscombe et Ducrot 1983, p. 36].

1.3 La théorie des topoï

La théorie des topoï³ s'inscrit dans la continuité de l'ADL, à laquelle elle reprend nombre de principes tout en introduisant les principes aristotéliens des *topiques*.

1.3.1 Principes topiques de l'argumentation

Les principes de base préalables à la théorie topique de l'argumentation ont été formulés par Ducrot à l'occasion de son séminaire intitulé *Sémantique des langues naturelles* [Ducrot 1993], puis transcrits par Grandchamp dans sa thèse [Grandchamp 1996], document à partir duquel nous les restituons. Cinq principes généraux sous-tendent l'argumentation :

- « décrire une phrase, c'est indiquer comment ses énoncés peuvent être employés dans les enchaînements discursifs » [Grandchamp 1996, p. 35] ; la description d'une phrase tient compte des enchaînements argumentatifs, réels ou virtuels, mais possibles, à partir des emplois de cette phrase en situation d'énonciation ;
- « les enchaînements discursifs sont fondés sur des principes généraux appelés topoï » [Grandchamp 1996, p. 35] ; les topoï sont convoqués *par* et *dans* le discours. Par exemple, le topos associant *richesse* et *bonheur* est susceptible de sous-tendre un enchaînement discursif comme :

Ex : *Il est riche, il a de la chance.*

- « décrire une phrase, c'est donner des « indications » sur les topoï permettant d'enchaîner ses énoncés dans le discours » [Grandchamp 1996, p. 35] ; cette description est faite hors contexte ;
- « qualifier linguistiquement un objet (être, objet concret, situation...) au moyen d'une phrase, c'est « convoquer » à propos de cet objet certains topoï permettant de construire des enchaînements discursifs à partir de cette phrase » [Grandchamp 1996, p. 35] ;

Ex : *Il est riche* (description), *donc travailleur* (orientation).

- « la convocation d'un topos peut être plus ou moins forte » [Grandchamp 1996, p. 35] ;

Ex : *Il est très riche, donc vraiment heureux.* Cet énoncé convoque le même topos *richesse* et *bonheur*, mais intensifié discursivement.

Ces principes déterminent un certain nombre de propriétés concernant les topoï.

³ Pour une présentation exhaustive des problématiques liées aux *topoi*, consulter l'ensemble des articles réunis dans [Anscombe 1995c].

1.3.2 Propriétés des topoï

La phrase se laisse donc décomposer en un enchaînement de deux segments, dont l'un est présenté comme justifiant l'autre, sous-tendus par un garant : le topos. Un topos est un principe argumentatif lié au mot et qui commande la façon dont on peut enchaîner à partir d'une phrase contenant ce mot. Les caractéristiques du topos sont les suivantes :

- le topos est une croyance considérée comme commune à une certaine collectivité dont font partie le locuteur et son interlocuteur ; elle est donc supposée partagée ;
- le topos est donné comme général, il vaut pour une multitude de situations discursives différentes, en tout cas distinctes de celle particulière dans laquelle le discours l'utilise. Général, mais pas universel : le topos est vrai dans la plupart des cas, mais pas invalidable ;
- le topos est graduel, il met en relation deux prédicats graduels.

Ex : *température de l'air et agrément de la promenade*

La relation qui unit les deux prédicats est graduelle (ou scalaire) elle aussi : on peut avoir plus ou moins de plaisir à aller se promener quand il fait bon dehors.

Le topos est donc un lieu commun qui autorise l'association de deux représentations mentales et permet leur expression linguistique par un lien de causalité donné comme naturel. La conséquence de ce postulat est que le topos peut apparaître sous deux formes topiques : une forme *concordante* ou une forme *discordante*.

Un topos est dit concordant s'il fixe pour ses deux prédicats graduels le même sens de parcours. Il y a deux formes topiques possibles : (+ P, + Q) et (- P, - Q).

Ex : + *travail*, + *réussite*

Ex : - *travail*, - *réussite*

Un topos est discordant quand les sens de parcours des deux échelles sont différents. Deux formes topiques sont également possibles : (+ P, - Q) et (- P, + Q).

Ex : + *économe*, - *dépenses*

Ex : - *économe*, + *dépenses*

Dans ce modèle topique de l'argumentation, la signification des unités lexicales est constituée uniquement de « paquets de topoï », sans que soit postulée l'existence d'un noyau dénotatif de type référentiel auquel les topoï seraient connotés. « Ainsi on introduisait dans la sémantique du mot travail des « croyances » comme « le travail fatigue », ou « le travail fait réussir », croyances censées fonder les argumentations possibles à partir des énoncés où ce mot intervient. » [Carel et Ducrot 1999, p. 8].

1.3.3 Topoï intrinsèques et extrinsèques

Les topoï apparaissent comme des garants d'enchaînements argumentatifs, au sens où ils justifient un raisonnement. Dans ce modèle, les énoncés jouent le rôle de propositions, les topoï, celui d'axiome et les règles que l'on peut énoncer pour le contrôle ou le chaînage des topoï, celui de règles d'inférence. Les topoï sont donc à considérer comme des unités

participant de la signification des mots et également comme des règles inférentielles. Dans cette perspective, les deux méta-prédicats liés par les topoï sont indissociables de l'emploi d'un mot donné, nécessairement associé en contexte pour sous-tendre un raisonnement précis. Dès lors, la prédictibilité des énoncés devient triviale. Pour pallier ce problème, la théorie des topoï évolue en distinguant entre les *topoï intrinsèques* et les *topoï extrinsèques*.

Les topoï intrinsèques fondent la signification de l'unité lexicale ou de la séquence de mots. Ils sont mis en jeu dans les énoncés pour lesquels le second membre ne fait qu'explicitier ce qui est déjà potentiellement contenu dans le premier.

Le niveau lexical réfère à la signification lexicale d'un mot.

Ex : *Pierre est riche, il peut donc s'offrir n'importe quoi.* [Anscombe 1995b]

Cet énoncé repose sur un topos intrinsèque qui postule que l'on ne peut parler de richesse sans évoquer le pouvoir d'achat.

Le niveau syntagmatique désigne une argumentation fondée sur la signification d'un groupe de mots.

Ex : *C'est une sale guerre. Il faut donc punir les responsables.* [Galatanu 1999a]

Cet énoncé repose sur un topos intrinsèque dont la conclusion résulte du syntagme *sale guerre*.

Les topoï extrinsèques sont utilisés pour « fonder les enchaînements discursifs » [Anscombe 1995b, p. 57]. Ils justifient certains enchaînements, non présents initialement dans la signification des mots employés. Deux types de conclusions sont donc possibles.

Ex : *Pierre est riche. Il est donc avare.*

Ex : *Pierre est riche. Il n'est donc pas avare.*

Ces énoncés établissent un lien argumentatif entre la richesse et l'avarice.

Les topoï extrinsèques peuvent être *culturels* ou *inédits*.

L'aspect culturel réfère à la nature du topos en jeu.

Ex : *Albert est un génie, il est donc invivable.* [Anscombe 1995b]

L'aspect inédit désigne un topos proposé par le sujet parlant et présenté comme déjà partagé par son destinataire.

Ex : *Albert est un génie, il n'aime donc pas les fleurs.* [Galatanu 2002b]

Si l'introduction des topoï extrinsèques permet à la langue de recouvrir sa dimension créative, notamment par les aspects culturel et inédit, le concept de topoï intrinsèques pose un nouveau problème, en lien avec la stabilité des représentations topiques d'une unité lexicale. Par exemple, comment interpréter un énoncé qui nierait telle ou telle propriété du mot *travail*, propriétés dont l'attribution fait partie du sens de ce mot ?

Ex : *Le travail repose.*

Cet énoncé contredit un topos qui semble bien faire partie des croyances communes, à savoir *le travail fatigue*. Dans ce cas, la sémantique argumentative, constatant la contradiction interne à l'énoncé, serait contrainte de statuer sur son absurdité. Naturellement, une telle analyse n'est pas satisfaisante.

Pour contourner cet écueil, Ducrot recourt à la notion de polyphonie. « L'idée de base est qu'en utilisant un mot un locuteur convoque des énonciateurs, dits lexicaux, qui posent les topoï inclus dans la signification de ce mot. Ainsi, en employant le mot *travail*, on introduit dans son discours une voix posant que le travail fatigue, qu'il donne des chances de réussite, qu'il mérite salaire, etc. Ces énonciateurs permettent de circonscrire un certain nombre d'objets (dans notre exemple, des activités), ceux auxquels on attribue généralement les propriétés prévues par les topoï du mot. Si maintenant le locuteur soutient que le travail

repose, il fait intervenir un nouvel énonciateur, auquel il s'identifie : cet énonciateur prédique, à propos des activités repérées grâce aux topoï du mot, une propriété, « reposer », contraire à l'un de ces topoï. » [Carel et Ducrot 1999, p. 7-8]. Toutefois, l'intervention d'un nouvel énonciateur pour expliquer le sémantisme des énoncés paradoxaux pose un problème de cohérence interne au modèle de la sémantique argumentative. De fait, une telle solution « repose sur une conception référentialiste de la prédication (« prédiquer, c'est attribuer une propriété à un objet ») : ainsi le dernier énonciateur, dans l'exemple du travail qui repose, déclarerait reposantes les activités repérées à partir des croyances exprimées par le premier. C'est le référent qui met en rapport les deux énonciateurs. Or la sémantique que nous voulons construire se veut purement discursive et ne peut pas faire intervenir, dans ses descriptions, des objets extra-linguistiques, notamment la réalité dont le discours est supposé parler » [Carel et Ducrot 1999, p. 8].

1.4 Les blocs d'argumentations

1.4.1 Préceptes de la théorie des blocs d'argumentations

La sémantique des blocs argumentatifs – SBA – est développée par Carel et Ducrot depuis 1999 [Carel et Ducrot 1999 ; Ducrot et Carel 1999]. Elle représente une nouvelle mouture de la théorie de l'ADL, tout en évitant les notions de topos et de forme topique.

La SBA part du principe que les discours du type argument – conclusion en DONC – DC – (car, si...) et en POURTANT – PT – (même si, malgré...) sont non pas inférentiels mais argumentatifs. Par exemple, l'argumentativité des discours en DC tient à une certaine interdépendance sémantique liée à DC, des mots qui constituent ces discours.

Ex : *La question des promotions pose problème, DC Paul propose de reporter sa discussion.*

Ex : *La question des promotions pose problème, PT Paul propose de la résoudre.*

Carel et Ducrot constatent que les liens établis par ces énoncés entre leurs mots respectifs sont apparentés. Pour rendre compte de cela, ils créent la notion de *bloc*, dans lequel s'inscrivent aussi bien les éléments de *problème DC report* que ceux de *problème PT NON-report*. Dans cette perspective, les deux énoncés ci-dessus appartiennent au même bloc, sans que le discours en DC n'ait de priorité sur le discours en PT : « Le discours en PT ne repose sur aucun discours en DC préalable qu'il concéderait. De manière plus générale, tout bloc a ainsi deux aspects argumentatifs : un aspect normatif, P DC Q, regroupant les discours en DC, et un aspect transgressif, P PT NON-Q, regroupant les discours en PT. Nous dirons parfois que les enchaînements réalisent les aspects auxquels ils appartiennent. » [Carel et Ducrot 1999, p. 12]. En définissant le caractère argumentatif d'un enchaînement par l'interdépendance sémantique de ses deux segments, Carel et Ducrot ne convoquent plus les concepts de topoï et formes topiques. Les deux types d'enchaînements en DC et PT constituent des unités sémantiques fondamentales. De ce fait, une argumentation peut être associée à une expression de deux façons : *externe* et *interne*. Si l'expression étudiée intervient dans l'un des enchaînements de l'aspect argumentatif qui lui est associé, cet aspect relève de l'argumentation externe, sinon cet aspect relève de l'argumentation interne.

1.4.2 Argumentation externe

Chronologiquement, en décrivant tout énoncé comme un argument pour certaines conclusions, Anscombe et Ducrot [Anscombe 1995c] ont analysé les premiers le fonctionnement de l'argumentation externe, sans toutefois employer ce terme. Chaque énoncé était associé à des instructions quant à ses suites possibles, les instructions étant sous-tendues par les topoï des enchaînements en DC, dont il pouvait être l'argument. Carel et Ducrot [Carel et Ducrot 1999] reprennent cette idée, sans mobiliser les concepts de topoï et de formes topiques. Ils associent à chaque énoncé les aspects des enchaînements en DC ou en PT dont il peut être un segment matériel.

Ex : *Paul est prudent, mais Marie ne lui fait pas confiance.*

Ce discours articule deux énoncés :

- Paul est prudent ;
- Marie ne lui fait pas confiance.

L'exemple ci-dessus articule les deux énoncés précédents comme les expressions de deux points de vue opposés. C'est ce qui fait dire aux auteurs que ces énoncés ont le statut d'unités sémantiques indépendantes, au sens où une telle articulation les associe à des discours argumentatifs. L'énoncé *Paul est prudent* évoque dans ce discours l'enchaînement normatif en DC – *Paul est prudent, DC Marie lui fait confiance*. Dans cette mesure, l'aspect « confiance » est dit appartenir à l'argumentation externe de « prudence ». De même, l'énoncé *Marie ne lui fait pas confiance* évoque dans ce discours l'enchaînement transgressif en PT – *Paul est prudent, PT Marie ne lui fait pas confiance*. Dans la mesure où *Marie ne lui fait pas confiance* résume l'enchaînement *prudent PT non-confiance*, cet aspect est déclaré faire partie de l'argumentation externe de *non-confiance*. Carel conclut à propos de la notion d'argumentation externe qu'elle « généralise la notion de potentiel conclusif utilisée par Anscombe et Ducrot puisque nous associons à certains énoncés non seulement leurs suites possibles en DC mais aussi des enchaînements en PT et également des enchaînements dont ils sont le second terme » [Carel et Ducrot 1999, p. 14].

1.4.3 Argumentation interne

L'introduction de l'argumentation interne constitue un nouveau type d'élargissement dont l'idée est d'associer certains énoncés à des enchaînements qui leur sont intérieurs.

Ex : *Ce bon étudiant a tout de même réussi.*

Ce discours qualifié de « curieux » par Carel articule deux énoncés :

- c'est un bon étudiant ;
- il a réussi.

Ces deux énoncés s'articulent sous l'aspect transgressif – *bon PT réussir* –, qui à l'intérieur de l'exemple relie le sujet grammatical *bon étudiant* et le prédicat *réussir*. Toutefois, comme l'exemple ne s'identifie pas directement aux enchaînements appartenant à *bon étudiant PT NON-réussir*, cet aspect relève de son argumentation interne.

Les énoncés ne sont pas les seules entités pourvues d’une argumentation externe et interne ; on peut aussi associer de telles argumentations aux syntagmes nominaux et aux mots.

Ex : *Paul est prudent, il n’aura donc pas d’accident.*

Cet énoncé évoque l’aspect argumentatif *prudent DC sécurité*. Dans la mesure où *prudent* intervient à l’intérieur de cet enchaînement, cet aspect relève de l’argumentation externe de *prudent*. En revanche, *prudent* provoque également des enchaînements comme :

Ex : *S’il y a du danger, Paul prendra des précautions.*

Cet énoncé évoque l’aspect argumentatif *danger DC précaution*. Dans la mesure où *prudent* n’intervient pas dans cet enchaînement, *danger DC précaution* est qualifié d’intérieur à *prudent* et relève par conséquent de son argumentation interne.

La présentation des différentes évolutions de la sémantique argumentative a permis, d’une part, de mettre en évidence les préceptes stables du modèle, à commencer par la dimension essentiellement discursive (argumentative) de la signification, et, d’autre part, de comprendre l’émergence des notions d’argumentation externe *vs* interne. Appréhender les progressions de ce modèle constitue un préliminaire nécessaire à la présentation et à la compréhension de la sémantique des possibles argumentatifs, puisque cette théorie s’inscrit dans la filiation directe de la sémantique des blocs, à laquelle elle emprunte les notions d’argumentation externe et interne. Ces deux concepts y sont centraux, dans la mesure où ils fondent la description de la signification lexicale, en la structurant de façon argumentative.

2 La sémantique des possibles argumentatifs

La sémantique des possibles argumentatifs – SPA –, développée par Galatanu depuis 1999 [Galatanu 1999a, 2002a, 2002b, 2004, 2005e, 2005f, 2005g, 2006], est un modèle de description-construction-représentation (désormais description) des significations lexicales susceptible de rendre compte du *cinétisme* des significations – entendu comme évolution –, fondé sur des mécanismes sémantico-discursifs et pragmatico-discursifs d’activation du potentiel argumentatif des mots en situation de discours. Cette description de la signification lexicale est donc basée simultanément sur des informations sémantiques (internes à la langue) et discursives (externes à la langue) ; c’est pourquoi elle consiste bien souvent en une reconstruction de la signification lexicale d’un mot à partir des hypothèses émises dans et par l’interprétation du sens de différentes occurrences d’emploi du mot concerné.

Nous présentons les fondements de la SPA, ses principes de description de la signification lexicale et les applications qui en découlent, en interaction nécessaire avec l’analyse linguistique du discours – ALD.

2.1 Filiations et postulats

2.1.1 Filiations logico-structuraliste et historico-psychologique

Historiquement, deux grandes traditions linguistiques s’opposent sur la conception de la signification – du sens, l’une héritée de la sémantique logique et structuraliste, où « le sens est d’abord défini comme une entité objective indépendante de l’esprit individuel : il relève de

la structure supra-individuelle du langage, des conditions de vérité dans les mondes possibles » [Geeraerts 1991, p. 45], l'autre héritée de la sémantique historico-psychologique et cognitive, où la conception du sens tient compte « de l'enracinement subjectif et individuel dans le vécu subjectif et individuel du langage » [Geeraerts 1991, p. 45]. L'idée sous-jacente à ce rapprochement repose sur un processus de « stabilisation du monde dans et par la langue », processus nécessairement subjectif et intersubjectif [Kerbrat-Orecchioni 1980], au sens où il est inévitablement variable dans l'espace individuel et social, donc marqué d'un flou cinétique.

En adhérant à ce processus de « stabilisation subjectif et intersubjectif du monde dans et par la langue », la SPA réconcilie ces deux traditions linguistiques. Plus précisément, la SPA associe, d'une part, une réflexion sur la proposition de Putnam [Putnam 1975, 1990, 1994] de décrire la signification lexicale des mots en termes de noyau (traits de caractérisation) et de stéréotypes (associés durablement aux mots) et, d'autre part, une conception de la signification orientée vers le potentiel discursif des mots, potentiel argumentatif qui fait l'objet privilégié de la sémantique argumentative dans ses différents développements : théorie de l'argumentation dans la langue [Anscombe et Ducrot 1983], théorie des topoï [Anscombe 1995c ; Ducrot 1995b], puis théorie des blocs d'argumentations [Carel 1995 ; Carel et Ducrot 1999, Ducrot et Carel 1999]. Cette filiation posée, le modèle SPA assigne à la signification lexicale un aspect tridimensionnel, à la fois dénotatif, argumentatif et par répercussion associatif.

2.1.2 Aspect tridimensionnel de la signification

2.1.2.1 Une approche dénotative et référentielle

« Dans une vaste série de cas, nos conceptualisations ou notre modèle mental du monde est largement identique d'un individu à l'autre, ce qui forme une sorte de socle pour une intercompréhension réussie. » [Kleiber 1999, pp. 21-22]. Dans le modèle SPA, la signification lexicale se présente en effet comme un processus d'abstraction concrétisé par une unité lexicale et visant une stabilisation du monde dans la langue. Aussi, la description de la signification lexicale des mots en termes de noyau et de stéréotypes [Putnam 1975, 1990, 1994] mobilise préalablement une conception référentielle de l'acte de dénomination. Dans la lignée des travaux de Kleiber [Kleiber 1984, 1999], le processus de dénomination correspond, d'une part, à « l'institution entre un objet et un signe X, d'une association référentielle durable » [Kleiber 1984, p. 79] et, d'autre part, à l'activation de cette association par l'acte discursif d'un locuteur. Le processus de dénomination a pour conséquence l'acquisition par les locuteurs d'une compétence référentielle correspondant à la mémorisation et au codage d'unités simples (**Ex** : *colère, guerre*) ou de lexies complexes (**Ex** : *colère noire, état de droit*) [Pottier 1974, p. 266].

La description de la signification lexicale telle qu'elle est formulée dans la SPA en trois strates (*cf. infra* : noyau, stéréotypes et possibles argumentatifs) permet d'envisager trois types d'actes de dénomination. Les deux premiers correspondent à l'aspect « référentiel durable » et concernent de ce fait le noyau de la signification lexicale, tandis que le dernier

correspond à « l'activation discursive » et concerne par conséquent les stéréotypes de la signification lexicale.

Le premier acte correspond à la fixation du lien référentiel par un acte de dénomination métalinguistique, « qui débouche sur la relation de signification descriptive et qui est actualisée par des phrases de dénomination [...] à caractère analytique, valables pour tous les exemplaires de la catégorie d'objets à laquelle renvoie le signe X et valant pour tous les locuteurs qui utilisent le signe » [Galatanu 2006, p. 6, à paraître]. Cette fixation, qu'il s'agisse d'un acte de dénomination ou d'une habitude associative, correspond dans le modèle SPA aux propriétés intrinsèques essentielles présentes dans le noyau.

Le deuxième acte correspond à l'actualisation de la relation de dénomination, qui convoque des propriétés intrinsèques essentielles de l'objet X, comme étant des traits définitoires du noyau.

Le troisième acte correspond à l'acte discursif de prise en charge de la dénomination par le locuteur, propre à générer des propriétés intrinsèques à la signification lexicale accidentelles. Cet acte équivaut à une rectification du lien référentiel débouchant sur une relation de signification nouvelle, partiellement reconnue par les locuteurs qui ont une compétence sémantique couvrant le noyau et les stéréotypes.

2.1.2.2 Une approche argumentative et encyclopédique

La SPA trouve une partie de ses fondements théoriques en sémantique argumentative ; elle croit à une organisation argumentative des représentations du monde dans et par la signification lexicale. Dans cette perspective, l'argumentation est définie comme « un acte sous-tendu par deux opérations mentales : une opération d'association de deux représentations du monde dans « un bloc de signifiant » (de signification) et une opération de sélection qui permet de poser un lien « naturel » entre ces deux représentations (cause-effet, intention-moyen, phénomène-symptôme...) » [Galatanu 1999b, p. 47]. L'acte discursif est également « essentiellement argumentatif » : « Tout énoncé a une dimension argumentative, soit, de manière explicite, transparente, par son appartenance à une séquence argumentative (énoncé-argument – énoncé-conclusion), soit par sa visée argumentative, sémantico-discursive (inscrite dans les « possibles argumentatifs » de sa signification), ou pragmatico-discursive (s'appuyant sur des informations contextuelles). » [Galatanu 1999b, p. 49].

L'aspect « explicite » réfère aux argumentations dites séquentielles, dont la transparence est absente des argumentations par visées.

Ex : *Conscient de la richesse des initiatives locales et de la nécessité de pouvoir encourager ces initiatives, en 1994 le ministre de l'époque a créé un dispositif de repérage et de valorisation des innovations.* (Exemple extrait de [Galatanu 2005d].)

L'argumentation est dite séquentielle, car l'énoncé comporte un argument *nécessité d'encourager les initiatives locales* et une conclusion (*Donc*) *création d'un dispositif de repérage et de valorisation des innovations*.

Ex : *Je rappellerai tout à l'heure que l'innovation s'était engagée, dès l'origine en 1994, dans le cadre de la déconcentration.* (Exemple extrait de [Galatanu 2005d].)

L'argumentation est dite par visée (argumentative) lorsqu'un énoncé « [autorise] de par sa signification ou par inférence pragmatique, à partir du contexte, un certain type de conclusions et [rend] d'autres conclusions peu probables » [Galatanu 2002b, p. 95]. L'énoncé comporte un argument *engagement de l'innovation dès l'origine*, mais pas de conclusion ; celle-ci est laissée à la charge du destinataire.

Les aspects sémantico-discursifs et pragmatico-discursifs réfèrent à deux phénomènes discursifs différents. Les phénomènes sémantico-discursifs s'appuient sur un degré de stabilité très élevée de l'association des représentations sémantiques, association faisant l'objet d'une convention linguistique présente dans la signification des mots. Tandis que les phénomènes pragmatico-discursifs correspondent à des associations de représentations proposées par l'acte de discours, dans la singularité de son contexte.

Ainsi, lorsque les conditions de bonne réalisation d'un acte de langage sont remplies, l'interprétation du sens des énoncés produits a nécessairement une dimension argumentative première. Cette dimension argumentative peut être représentée par un *ensemble ouvert de séquences discursives* que l'énoncé autorise et produit. Ces *ensembles ouverts* sont ouverts dans ce sens qu'il serait impossible d'identifier avec certitude des limites rigides à ces ensembles dans une communauté linguistique à un moment donné de l'évolution de sa langue. La SPA postule pouvoir rendre compte de cet ensemble ouvert d'enchaînements argumentatifs au niveau de la description même de la signification lexicale, tout en articulant cette inscription du potentiel argumentatif à la lecture descriptive du monde représenté et stabilisé par les mots. Cette opération est possible car « si tout énoncé actualise un potentiel argumentatif, cela revient à dire que tout énoncé contient une évaluation de par le lien argumentatif posé et que tout lexème est potentiellement porteur d'une évaluation » [Galatanu 2005e, à paraître]. La flexibilité de la structure interne de la signification lexicale, fruit des changements de sens, fait de la SPA un modèle encyclopédique : « Tous les aspects de notre connaissance de l'entité en jeu contribuent au sens de l'expression qui la désigne. » [Galatanu 2005d, à paraître]. La dimension sémantico-discursive et pragmatico-discursive de la signification lexicale, tous les aspects de notre connaissance participent à sa construction-reconstruction discursive.

2.1.2.3 Une approche associative et holistique

Par répercussion, ces dimensions fondamentalement dénotative *et* argumentative confèrent à la signification lexicale un caractère associatif et holistique⁴.

La SPA appréhende le langage en tant qu'outil argumentatif permettant, au niveau cognitif, la construction des représentations du monde perçues et modélisées par la langue (au travers des significations) [Kleiber 1999, pp. 27-34] et, au niveau énonciatif, l'expression de soi à travers la communication du monde représenté. En ce sens, la SPA est une approche de la signification lexicale associative, puisqu'elle cherche à rendre compte de l'ancrage dénotatif de celle-ci et de son potentiel argumentatif. Concrètement, ce caractère associatif est le fruit d'une conjugaison de la partie stable de la signification (le noyau), composée de propriétés essentielles du concept, et d'une partie argumentative de la signification (les stéréotypes), décrite en termes de blocs d'argumentations. Cette conjugaison est synergique, au sens où les associations entre le noyau et les stéréotypes génèrent des représentations sémantiques relativement stables, mais formant des ensembles ouverts [Galatanu 2005g]. On retrouve ici la dimension encyclopédique du modèle, sur laquelle repose l'aspect holistique de la signification. Ce dispositif d'ensembles ouverts invite à l'exhaustivité des stéréotypes. D'une part, cette exhaustivité revêt un caractère prédictible ; s'il est impossible de dénombrer

⁴ Appliqué à la signification lexicale, *holistique* signifie « objet constituant un tout, représentant une totalité ».

l'ensemble (cinétique) des stéréotypes attachés à une unité lexicale, en revanche, il est possible de statuer sur leur plausibilité. D'autre part, cette exhaustivité revêt un caractère culturel ; les associations entre le noyau et les stéréotypes intègrent les variations dues à la diversité et à l'évolution des représentations culturelles. Parce que ces associations ont la capacité d'inscrire perpétuellement de nouveaux éléments relevant du contexte culturel et contextuel, elles sont susceptibles de rendre compte en permanence d'une totalité (individuelle, culturelle...), bien que la réalité physique de l'acte énonciatif ne permette d'en mobiliser que quelques-uns à la fois.

2.2 Représentation de la signification lexicale

2.2.1 Les conditions nécessaires

Compte tenu des postulats précédemment énoncés, la représentation sémantique proposée par la SPA pour la signification lexicale est contrainte par quatre conditions devant être simultanément remplies. Cette représentation se doit de :

- comporter une dimension descriptive de la signification lexicale, préservant ainsi le recours à une lecture de la *forme logique* de la phrase énoncée [Sperber 1996] ;
- faire apparaître à la fois la partie « stable » de la signification et sa partie évolutive, que le discours charge et/ou décharge de valeurs axiologiques (**Ex** : *innovation* [Galatanu 2005b]) ;
- rendre compte du statut d'ensemble ouvert des éléments de signification évolutifs, appelés « stéréotypes » (au sens de Putnam et Anscombe) ;
- rendre compte du potentiel discursif (argumentatif) de la signification lexicale.

Aussi, la SPA décrit la signification lexicale d'un mot comme étant composée de trois strates : un noyau, des stéréotypes et des possibles argumentatifs.

2.2.2 Le noyau et les stéréotypes

Le noyau⁵ et les stéréotypes sont des entités sémantiques.

Le noyau est composé de traits nécessaires de catégorisation – TNC – et de prédicats abstraits. Les TNC rassemblent trois types d'informations : la nature grammaticale du mot (**Ex** : nom, verbe...), ses caractéristiques modales⁶ (**Ex** : épistémique, axiologique...) et son orientation axiologique (positive ou négative). Les prédicats abstraits correspondent à des

⁵ Le travail sur la description du noyau a connu plusieurs étapes : une réflexion sur la possibilité d'élaborer une représentation en termes de « primitifs sémantiques », une réflexion sur la représentation possible en termes de structures prédictives ou structures profondes sémantico-syntaxiques, une réflexion sur la représentation des différentes parties du discours, en particulier les catégories nominales et verbales [Galatanu 2003a, 2003b, 2005a].

⁶ Pour une typologie complète des zones modales, consulter [Galatanu 2002c, pp. 17-32].

primitifs sémantiques invariants⁷ [Putnam 1994], représentant la partie stable de la signification⁸ lexicale d'un mot.

Putnam définit le stéréotype comme une partie de la signification d'un mot ; en particulier celle qui comprend les traits associés fréquemment au contenu d'un mot, y compris les marqueurs syntaxiques et sémantiques (traits nécessaires pour la catégorisation des « objets » auxquels le mot réfère). Dans le modèle SPA, les stéréotypes correspondent à la déclinaison des prédicats abstraits du noyau par association avec d'autres invariants (représentations), dans des blocs de signification argumentatifs posés comme naturels (**Ex** : cause-effet, symptôme-phénomènes, but-moyen...). Ces ensembles ouverts constituent des blocs d'argumentations internes à la signification lexicale. Un bloc d'argumentation interne représente une association de prédicats appartenant à la signification d'une entité lexicale dans une structure argumentative [Carel et Ducrot 1999].

2.2.3 Les possibles argumentatifs

Les possibles argumentatifs – PA – sont des entités discursives. Cette troisième strate, inexistante dans la description de Putnam, correspond à l'association réelle ou virtuelle, mais prévisible, de deux unités lexicales liées par une relation d'argumentation et calculable sur la base des enchaînements linguistiques que la signification lexicale est susceptible de générer en contexte. Plus clairement, il s'agit de la concrétisation dans le discours de l'association du mot avec l'un des éléments de ses stéréotypes. Le caractère réel ou virtuel d'un PA est fonction de la nature séquentielle ou par visée de l'argumentation énoncée. L'ensemble formé par les stéréotypes fonctionne donc comme un dispositif de génération de discours potentiels. Ils sont activés en cotexte et/ou en contexte⁹ par un phénomène de « contamination discursive ». On parle de blocs d'argumentations externes à la signification lexicale, puisque consommés dans le discours. Un bloc d'argumentation externe représente l'association, dans une structure argumentative, du mot dont on construit la signification lexicale avec l'un des prédicats identifiés comme faisant partie de son argumentation interne [Carel et Ducrot 1999].

Pour exemplifier ces trois strates inhérentes à la construction de la SL figure ci-dessous la représentation sous forme de tableau du mot *vertu* [Galatanu 2005a, p. 162-163].

⁷ Voir les analyses de Putnam concernant les mots *arbre* et *eau* [Putnam 1975, 1990, 1994].

⁸ Pour Putnam, « connaître la signification d'un mot peut vouloir dire : (a) savoir la traduire, ou (b) savoir ce qu'il désigne au sens de : avoir la capacité d'énoncer explicitement ce qu'est la dénotation (autrement qu'en utilisant le mot lui-même), ou (c) avoir une connaissance tacite de sa signification, au sens d'être capable d'utiliser le mot dans son discours. Le seul sens dans lequel le locuteur moyen du langage connaît la signification de la plupart des mots est le sens (c) » [Putnam 1990, p. 74].

⁹ Le cotexte représente l'environnement textuel, alors que le contexte représente l'environnement extra-linguistique (contexte social, contexte culturel, contexte situationnel du discours...).

VERTU			
TRAITS NECESSAIRES DE CATEGORISATION (TNC)			
TYPE GRAMMATICAL		ZONE MODALE	ORIENTATION AXIOLOGIQUE
Nom – Féminin		Affectif – Moral	Axiologique – Positif
ENTITES SEMANTIQUES			ENTITES DISCURSIVES
NOYAU		STEREOTYPES	POSSIBLES ARGUMENTATIFS
1	Processus : devoir bien faire, vouloir bien faire	Donc devoir ne pas mentir, travailler... Donc ne pas tuer, défendre les pauvres...	V. donc non menteur
2	Processus : vouloir bien faire	Donc ne pas convoiter la femme de son voisin... Donc vouloir aider autrui...	V. donc respectueux
3	Résultat : bien faire	Donc ne pas tuer... Donc ne pas mentir... Donc travailler...	V. donc travailleur

Tableau 1 : Signification lexicale de vertu

De par le lien argumentatif posé, les PA sont par essence évaluatifs. Ils s'organisent autour de deux faisceaux orientés respectivement vers l'un ou l'autre des pôles axiologiques, positif ou négatif. Les pôles axiologiques se définissent par rapport à la zone modale de l'axiologique, qui « recouvre une zone sémantique, définie par un postulat empirique, qui renvoie à l'idée de préférence, de rupture de l'indifférence (étymologiquement : axio = « ce qui vaut »), *id est* à une polarité : positif/négatif, bon/mauvais, bien/mal » [Galatanu 2000b, p. 94]. « Cette zone comporte des évaluations liées à des champs d'expérience humaine : esthétique (beau – laid), pragmatique (utile – inutile, efficace – inefficace), cognitif ou intellectuel (intéressant – inintéressant), hédonique – affectif (agréable – désagréable). » [Galatanu 2005a, p. 164]. L'orientation positive ou négative des PA est fonction de la contamination discursive, due soit à l'environnement sémantique, soit à l'environnement cotextuel et/ou contextuel.

2.2.3.1 Les valeurs axiologiques

Les valeurs axiologiques associées à un mot posent la question du niveau d'inscription de ces valeurs (modales) dans la signification lexicale de l'unité considérée. Ces valeurs axiologiques peuvent être soit positives, soit négatives, soit les deux. Lorsque la valeur d'un mot est soit positive, soit négative, le mot est dit axiologiquement *monovalent*.

Ex : *Beau, vertueux* sont monovalents positifs.

Ex : *Laid, idiot* sont monovalents négatifs.

Lorsque la valeur d'un mot est potentiellement positive et négative, le mot est dit axiologiquement *bivalent*.

Ex : *Guerre, grève* sont bivalents, ils conservent un potentiel positif et négatif inscrit aux niveaux du noyau et des stéréotypes.

De même que l'aspect argumentatif de la langue est inscrit à différents niveaux de représentation de la signification lexicale, les valeurs axiologiques peuvent être inscrites à différentes strates de la signification lexicale : au niveau du noyau, des stéréotypes ou encore des possibles argumentatifs.

Ex : *Intelligent* est monovalent, le pôle positif est inscrit dans son noyau.

Ex : *Mauvais* est monovalent, le pôle négatif est également inscrit dans son noyau.

Ex : *Démocratie* est monovalent, le pôle (moralement) positif est inscrit dans ses stéréotypes.

Ex : *Crime, vol* sont monovalents, le pôle négatif est inscrit dans leurs stéréotypes.

Ex : *Guerre, grève* sont bivalents, cette bivalence fait partie des PA de leurs stéréotypes. « L'un ou l'autre des pôles axiologiques (positifs ou négatifs) va être sélectionné et activé dans le discours par un processus de contamination avec les stéréotypes des mots de leur environnement, stabilisant ainsi une orientation argumentative : *sale guerre, guerre juste, guerre de défense*. » [Galatanu 2005a, p. 166].

Dans le cas où les mots sont axiologiquement marqués, c'est-à-dire que l'un des pôles positif ou négatif est inscrit dans leur noyau (**Ex :** *bien, mal, vertu*) ou dans leurs stéréotypes (**Ex :** *viol, égalité*), le potentiel axiologique des PA reste double, à la fois positif et négatif. En effet, l'association du mot avec un élément de son stéréotype peut prendre une orientation conforme ou opposée à l'orientation axiologique du stéréotype.

Ex : *Vertueux donc non menteur*. Orientation conforme au stéréotype.

Ex : *Vertueux pourtant menteur*. Orientation opposée au stéréotype.

C'est-à-dire que, contrairement au bloc d'argumentation interne (formé du noyau et des stéréotypes), qui ne peut connaître deux formes topiques à la fois, le bloc d'argumentation externe (formé du mot associé à l'un de ses stéréotypes) peut apparaître sous une forme normative ou transgressive.

Ex : *Vertueux donc respectueux*

Ex : *Vertueux pourtant irrespectueux*

2.2.3.2 Les mécanismes de l'axiologisation

Du point de vue sémantique, l'orientation axiologique du PA est fonction de la valeur axiologique positive ou négative de la signification lexicale du mot sur lequel repose l'enchaînement argumentatif.

Ex : *Elle est intelligente, donc douée à l'école*.

Dans cet exemple, on constate l'activation d'un PA axiologique positif porté par la signification lexicale de l'adjectif *intelligente*.

Ex : *Il est lâche, donc inutile de compter sur lui en cas d'attaque*.

À l'inverse, dans cet exemple, on observe l'activation d'un PA axiologique négatif porté par la signification lexicale de l'adjectif *lâche*.

Du point de vue co- et contextuel, l'orientation axiologique du PA est fonction des « caprices du discours », au sens où ils peuvent être « contaminés » par lui. Aussi, l'information co- et

contextuelle peut activer les PA dans des occurrences discursives, les renforcer, les affaiblir... Au total, six mécanismes ont été mis à jour : le renforcement, l'activation, la neutralisation, l'affaiblissement, l'interversion et la transgression.

Ex : Renforcement : *Elle est belle, donc mariée depuis très longtemps.*

Ex : Activation : *Elle est belle, donc elle plaît.*

Ex : Neutralisation : *Tous les ans il y a une grève générale.*

Ex : Affaiblissement : *Elle est belle, mais intelligente.*

Ex : Interversion : *C'est bon la honte.* (Publicité pour les biscuits Senoble)

Ex : Transgression : *Elle est belle, pourtant célibataire.*

Dans des travaux plus récents [Galatanu 2004, 2005d], un septième mécanisme dit « *stéréophagie* » s'est fait jour. Plus que l'interversion ou la transgression, la stéréophagie correspond simultanément à un phénomène de déconstruction-reconstruction de la signification lexicale. Concrètement, il s'agit du « développement de l'un des éléments du noyau jusqu'aux dernières conséquences, par la proposition de nouveaux stéréotypes, développement qui rend impossible ou paradoxale la présence dans la signification du mot, d'un autre élément de son noyau ou de ses stéréotypes » [Galatanu 2004, p. 222].

Ex : Stéréophagie (déconstruction-reconstruction) : *colère légitime*¹⁰

« La colère légitime rend paradoxal le stéréotype du mot colère, qui associe à un élément du noyau l'action violente, car cette dernière n'est pas légitime, ce qui justifie d'ailleurs le traitement de la colère, en d'autres temps, comme un « péché capital » et actuellement comme une quasi-pathologie. » [Galatanu 2004, p. 222]

Le mécanisme de stéréophagie participe donc du cinétisme de la signification lexicale, par la formation possible d'associations inédites au regard des représentations sémantiques stabilisées.

2.2.4 Le « monde quantique » de la signification

Dans les premiers développements de la SPA [Galatanu 1999a], la signification lexicale est décrite par analogie avec la structure quantique de l'atome dans le « monde quantique »¹¹. Malgré le caractère obsolète de certains parallèles, en particulier celui de « nuage topique », cette analogie reste valable et conserve un pouvoir éclairant.

Cette analogie repose :

- sur le noyau de la signification lexicale, comparable au noyau de l'atome, associé aux stéréotypes dans des structures argumentatives possibles ;
- sur la présence de stéréotypes (traits argumentatifs possibles), comparables aux électrons qui gravitent autour du noyau de l'atome. Ces électrons n'occupent pas de position précise ; ils sont diffus, au sens où leur présence en un lieu et à un moment donné repose sur un principe probabiliste ;
- sur la présence de PA « qui relie des éléments du stéréotype à d'autres représentations sémantiques (ou stéréotypes d'autres mots) et qui se superposent

¹⁰ Pour plus de clarté, consulter le tableau de la signification lexicale de *colère* proposé au chapitre 4.

¹¹ La physique quantique s'applique à l'échelle des atomes et des particules. Les objets quantiques ont des propriétés qui se caractérisent non pas par une valeur précise mais par un spectre de valeurs.

(comme les états des électrons, qui sont simultanément présents et absents) » [Galatanu 1999a, pp. 48-49]. À l'origine, ces PA étaient décrits en termes de « nuage topique » reliant les éléments du noyau à d'autres représentations sémantiques (électrons topiques).

Cette description des PA, jugée trop floue, a été précisée depuis. Les PA sont désormais décrits en termes de blocs d'argumentations externes [Galatanu 2005g]. De ce fait, ils récupèrent à la fois un déploiement argumentatif possible simultanément normatif et transgressif et un potentiel axiologique possible simultanément positif et négatif. Cette superposition d'états est comparable à la superposition d'états décrite dans la célèbre expérience de pensée de Schrödinger [Schrödinger 1935, 1992] concernant le chat, simultanément mort et vivant¹².

L'analogie entre la signification lexicale et le monde quantique va plus loin, puisque le modèle SPA récupère également le principe de *décohérence*. Les théoriciens quantiques affirment que cet état de superposition ne peut être maintenu qu'en l'absence d'interactions avec l'environnement. En effet, c'est l'environnement qui « déclenche » le choix entre les deux états (du chat mort ou vivant). La théorie de la décohérence postule que la rupture entre les deux états est provoquée par la première interaction physique qui se produit, annihilant ainsi l'état de superposition d'états, c'est-à-dire de cohérence. « Le phénomène de « décohérence », défini récemment par les chercheurs, explique pourquoi la physique quantique ne s'applique pas à notre échelle. Zurek [Zurek 1991] démontre que les multiples interactions des objets quantiques avec l'environnement entraînent une destruction rapide des interférences quantiques du système, la destruction de la cohérence des états quantiques, de sorte qu'il ne reste plus que des états simples. » [Galatanu 1999a, p. 49]. De même que, dès l'ouverture de la boîte, le chat recouvre un état physique mort ou vivant, dès la production d'un énoncé, le bloc d'argumentation formé par un PA instancie une forme normative ou transgressive, axiologiquement positive ou négative.

2.3 Applications et interactions

2.3.1 Considérations méthodologiques

Dans l'approche de la SPA, l'analyse sémantique des entités lexicales ne peut être qu'une construction de l'objet théorique « signification lexicale », à partir d'hypothèses externes sur le sens des occurrences du mot analysé en contexte. Cette première étape de l'analyse correspond, pour des raisons opérationnelles :

- soit à la collecte et à l'étude du/des discours lexicographique(s) ;
- soit à la collecte et à l'étude d'énoncés produits et/ou repérés dans des séquences de communication.

¹² « L'expérience « de pensée » de Schrödinger (1935) consiste à enfermer un chat dans une boîte avec un atome radioactif commandant un dispositif mortel. L'atome, qui est un système quantique, est dans une superposition d'états « désintégré – pas désintégré » et le système constitué du chat, du dispositif et de l'atome est aussi dans un état de superposition d'états « atome intact – chat vivant et atome désintégré – chat mort ». » [Galatanu 1999a, p. 49].

Naturellement, la seconde possibilité nécessitant la mise en place d'un protocole expérimental assez lourd (pour que la démarche soit fructueuse), la seconde solution est généralement préférée (quoiqu'une méthode mixte soit également envisageable). Dans cette perspective, l'analyse des définitions lexicographiques a pour objectif premier l'émergence des éléments stables du noyau de la signification du mot, qu'il importe de dissocier (s'il y a lieu) entre le processus et le résultat.

Ex : *innover* = introduire une chose nouvelle (selon les définitions des dictionnaires)

Le noyau d'*innover* peut donc être représenté par les blocs d'argumentations internes :

<action X (d'*innover*) donc résultat = chose nouvelle>;

<résultat = chose nouvelle parce qu'action X (d'*innover*)>¹³.

Les définitions étudiées concernant *innover* font toutes état « d'une situation établie ».

Après le noyau, la deuxième étape est celle de la construction des stéréotypes. De même que pour le noyau, il s'agit de puiser dans les informations présentes au sein de l'article lexicographique. Toujours concernant le mot *innover*, figurent ci-dessous les deux ensembles stéréotypiques correspondant aux deux noyaux présentés ci-dessus.

Ex : Action (de changer) ou résultat (changement) : donc évaluation positive/négative ; donc succès/échec social ; donc admiration/regard critique ; donc...

Ex : Situation établie à changer/changée : donc amélioration/désordre ; donc plus belle/plus laide ; donc plus efficace/moins efficace ; donc...

Les représentations proposées pour *innover* (sans prétendre au statut de formalisation) mettent en évidence la bivalence axiologique du mot, au niveau des stéréotypes.

Enfin, la troisième étape nécessaire à la mise en évidence du caractère cinétique des significations lexicales correspond au repérage des occurrences des mots étudiés dans un discours défini et délimité sur critères praxéologiques¹⁴. L'analyse des PA (des déploiements argumentatifs) des mots, sujets à la contamination discursive, va confirmer ou infirmer (c'est-à-dire renforcer, activer, affaiblir, neutraliser, intervertir...) le dispositif stéréotypique préalablement construit.

Ex : *Innover pour aller vers l'autre ; innover pour intéresser tout le temps ; innover pour répondre à un besoin ; innover pour ne pas reculer ; innover pour communiquer son expérience ; innover pour suivre la réalité ; innover pour réussir...*

C'est à partir de ces déploiements argumentatifs que l'on peut reconstruire les significations des mots, dont le cinétisme est posé d'emblée comme un *a priori* de l'approche SPA.

¹³ Tous les exemples concernant *innover* et *innovation* sont extraits de [Galatanu 2005b].

¹⁴ Dans le cas des exemples *innover*, le corpus est composé de l'ensemble des textes constituant les actes du séminaire « Pourquoi innover ? » qui s'est déroulé le 7 mai 2003 au lycée Pergaud, situé à Besançon.

2.3.2 Analyse linguistique du discours

Traditionnellement, les problématiques liées à l'analyse du discours s'articulent autour de deux objectifs :

- il s'agit en premier lieu d'identifier la spécificité du discours étudié, par la mise en évidence possible des éléments récurrents dans la pratique discursive abordée, des traits caractérisant l'identité discursive, ou encore d'une occurrence énonciative envisagée dans la singularité de l'acte de parole ;
- il s'agit également, à partir des résultats obtenus, de formuler des hypothèses interprétatives portant sur la pratique humaine qui sous-tend le discours étudié (analyse du discours au service des pratiques sociales).

Ces objectifs correspondent tous deux à une seule et même approche dite « classique » de l'analyse du discours¹⁵, qui inscrit son objet d'étude dans la praxéologie¹⁶ et s'intéresse de ce fait aux incidences identitaires (transformation de soi et élaboration d'une identité collective). Par son objet d'étude centré sur le cinétisme des significations lexicales, la SPA couvre deux types de phénomènes discursifs : sémantico-discursifs et pragmatico-discursifs. Or, que l'association des représentations sémantiques soit stabilisée par convention linguistique ou proposée par l'acte de discours dans la singularité de son contexte, étudier ces deux mécanismes implique en premier lieu de décrire le sens que le sujet interprétant peut attribuer aux énoncés-occurrences, pour ensuite re-construire la signification des entités linguistiques étudiées (*cf. supra*, considérations méthodologiques). De fait, le cinétisme des significations lexicales est fonction des occurrences discursives du sens des unités lexicales, susceptibles d'agir sur les systèmes de valeurs. Le sens est conçu comme « l'ensemble des représentations de soi et du monde et les systèmes de valeurs qui les accompagnent, représentations rendues publiques dans et par les énoncés qui forment le texte, résultat de l'activité discursive » [Galatanu 2002b, p. 96].

C'est pourquoi la SPA situe sa proposition théorique à l'interface entre la sémantique lexicale et l'analyse du discours, dans un « espace » qu'elle appelle analyse linguistique du discours – ALD – [Galatanu 1999b, 2000b, 2003a, 2004, 2005b, 2005d]. L'ALD se définit comme « l'analyse du discours orientée vers l'étude des mécanismes de production du sens habilités par le potentiel discursif des significations des entités linguistiques et plus particulièrement des unités lexicales » [Galatanu 2006, à paraître]. L'énoncé-occurrence constitue l'objet empirique de l'ALD, car il permet l'étude des mécanismes discursifs mis en œuvre dans la construction référentielle du texte. L'ALD constitue donc fondamentalement un modèle d'interprétation du sens des énoncés, dans la mesure où, d'une part, les représentations sémantiques stabilisées dans la signification lexicale sont éprouvées en situation d'énonciation (au sens où elles doivent être activées dans le discours) et où, d'autre part, d'autres représentations sémantiques sont susceptibles d'émerger du sens des énoncés et de modifier en retour les représentations stabilisées. En ce sens, l'articulation SPA-ALD fonctionne en interaction nécessaire.

¹⁵ Pour une présentation de l'analyse du discours dans la tradition française, consulter [Maingueneau 1996].

¹⁶ *Praxéologie* est employé au sens « d'étude de l'action humaine » (**Ex** : psychologie sociale, sociologie, sciences de l'éducation...).

Compte tenu de ses objectifs proprement sémantiques, l'ALD correspond à une approche dite « linguistique » de l'analyse du discours qui inscrit son objet d'étude dans une praxéologie nécessaire, mais orientée vers les incidences linguistiques (transformation du cinétisme des significations lexicales). Dans cette perspective, le discours est d'abord envisagé comme une voie d'accès aux connaissances des sujets et à leurs représentations. Par répercussion, ses représentations permettent l'accès aux identités des acteurs sociaux, qu'ils soient instances énonciatives ou qu'ils fassent partie seulement de l'univers référentiel du discours. Le discours reconstruit les systèmes de valeurs qui accompagnent les pratiques sociales, aussi, l'analyse des représentations sociales passe par l'analyse des discours portés par les acteurs sociaux. C'est pourquoi « le discours est également étudié de plus en plus dans sa dimension « d'acteur du monde » qui agit sur les états mentaux pour les conforter, les compléter, les enrichir, les insérer dans de nouveaux systèmes de connaissances et de valeurs, ou encore les déstabiliser, les déconstruire et les reconstruire » [Galatanu 2005b, à paraître].

Conclusion

Contrairement à la sémantique argumentative, la SPA postule que la langue générale et *a fortiori* la signification lexicale sont dénotatives, c'est-à-dire qu'elles résultent d'un besoin de dénomination rendu possible par une capacité de fixation référentielle ou d'habitudes associatives entre l'objet X et le signe linguistique x. Ce faisant, la SPA attribue à la signification lexicale l'existence d'un noyau contenant des informations stables sur lesquelles peuvent se construire les enchaînements argumentatifs, décrits en termes de blocs d'argumentations internes, les stéréotypes, et externes, les possibles argumentatifs. À la différence de leur représentation respective indépendante en sémantique argumentative, les blocs d'argumentations internes et externes sont articulés en SPA. La SPA est donc un modèle de description de la signification lexicale mixte, au sens où cette description intègre simultanément des informations sémantiques et discursives. C'est pourquoi, cette approche SPA permet au niveau sémantique, de mettre à jour des mécanismes explicatifs quant au cinétisme de la signification lexicale et par là de rendre compte du pouvoir qu'a le discours, de véhiculer des systèmes de valeur, de les transformer, mais aussi de les produire. Par voie de conséquence, au niveau discursif ALD, les analyses portent sur les mécanismes de renforcement, affaiblissement, interversion... des valeurs portées par les nominaux monovalents et sur les mécanismes d'activation des valeurs modales axiologiques pour les nominaux bivalents. Dans un troisième temps, du point de vue ALD, ce type d'analyse offre la possibilité de formuler des hypothèses interprétatives sur la fonction sociale du discours étudié et sur la reconstruction discursive des systèmes de valeurs.

La SPA-ALD inscrivant son objet d'étude dans une praxéologie nécessaire, le choix de textes permettant l'accès à un discours représentatif d'une communauté, d'un domaine, etc., constitue l'étape préliminaire à toute analyse.

Chapitre 3

Principes de constitution du corpus TAL(N)

Présentation

L'objectif de ce chapitre est double. Du point de vue théorique, il s'agit d'une part, en partant d'une définition courante de corpus, de parvenir à un concept scientifiquement exploitable de la notion de corpus électronique spécialisé. Cette démarche suppose parallèlement de choisir un cadre théorique approprié à la description d'une langue de spécialité. Il s'agit d'autre part de confronter et de sélectionner les différentes contraintes qui pèsent sur la constitution du corpus électronique spécialisé, compte tenu du cadre théorique adopté pour cette recherche. Du point de vue pratique, il faut d'abord appliquer les contraintes de constitution de corpus à la communauté TAL pressentie ; puis construire le corpus TAL(N) à proprement parler, c'est-à-dire présenter la méthodologie d'obtention du corpus et son résultat.

1 « corpus »

1.1 Le concept général

Lorsque l'on cherche la définition d'un mot, il est d'usage de consulter le dictionnaire. Nous recourrons donc au *Petit Robert* et au *Lexis Larousse*, parce qu'ils rassemblent sous une même entrée des définitions courantes (générales) du mot *corpus* et des définitions spécialisées au domaine linguistique.

Le Petit Robert (édition 2000) :

CORPUS : n. m. – 1863 ; « hostie » fin XII^e ; mot lat. « corps »

- DR. Recueil de pièces, de documents concernant une même discipline. *Corpus d'inscriptions latines et grecques*.
- (1961) LING. Ensemble fini d'énoncés réels réunis en vue de l'étude d'un phénomène linguistique. *Corpus écrit, oral. Corpus fermé, ouvert*, qui peut être augmenté.

Lexis Larousse (édition 1998) :

CORPUS : n. m. (lat. *corpus* [juris], recueil de droit ; 1863).

- Recueil concernant une même matière : il existe des corpus d'inscriptions latines et grecques.
- (1950) Ensemble de documents servant de base à la description ou à l'étude d'un phénomène.
- Ling. Ensemble fini d'énoncés, qui constitue le matériau destiné à la description linguistique.

Les trois définitions générales demeurent tout à fait imprécises. Quelles sortes de *documents-pièces* ? Quels types de *disciplines-matières* ? En outre, elles présentent une différence notoire, puisque l'une assigne au corpus une fonction descriptive ou analytique que l'on ne retrouve pas dans l'autre.

Dans son acception spécialisée au domaine linguistique, les données s'opacifient encore.

La première définition introduit :

- la notion de complétude – *ensemble fini d'énoncés* –, renvoyant au *corpus fermé* (par opposition à *ouvert*), c'est-à-dire au corpus dont le contenu ne peut être augmenté de données nouvelles ;
- la notion d'authenticité – *énoncés réels* –, renvoyant au *corpus attesté* (par opposition à *forgé*), c'est-à-dire au corpus composé d'énoncés réellement produits par des locuteurs.

Tandis que dans la seconde définition est de nouveau mentionnée une fonctionnalité descriptive dont on ignore si elle est indexée ou non sur une réalité de faits linguistiques.

En revanche, les définitions évoquent :

- la notion d'*énoncés*, référant possiblement à des données langagières écrites, *corpus écrit* (**Ex** : le corpus *Le Monde*, tiré du quotidien français), ou orales, *corpus oral* (**Ex** : le Corpus de référence du français parlé [Campione et al. 2005], réalisé par l'équipe DELIC dirigée par Jean Véronis) ;
- la notion de *recueil* ou d'*ensemble* de *documents-pièces-énoncés* ; elle renvoie donc indifféremment soit à des corpus du type *collection physique d'énoncés*, soit à des *corpus électroniques*.

Or il est tout à fait nécessaire de distinguer entre ces deux référents possibles car, conformément à leur nature, les méthodes d'analyses manuelles et logicielles employées pour les étudier diffèrent. L'adjectif *électronique* a son importance ; la création de corpus n'est pas une nouveauté en linguistique et, ce qui évolue actuellement, c'est précisément leur possible informatisation. La quantité de textes stockés est désormais considérable et leur exploitation informatique, nécessaire. D'ailleurs, la polysémie du terme *corpus* est propre à la langue française, puisque la définition anglaise du mot intègre indissociablement la dimension électronique – *machine readable* – exploitable par l'ordinateur [Williams 1999b].

Une *collection physique d'énoncés* est un ensemble limité de phrases ou encore une série d'énoncés, mais rarement un groupement de textes. Ces données sont couramment utilisées par les linguistes, toutes disciplines confondues (phonologie, syntaxe, sémantique...) pour conforter ou réfuter une position théorique.

À cet effet, les syntacticiens utilisent généralement des corpus de phrases forgées, construits pour éprouver l'hypothèse posée. Par exemple, Tellier [Tellier 1995, 2002], dans son ouvrage sur la syntaxe du français, propose le corpus suivant pour amorcer la validation de l'hypothèse générative selon laquelle certaines de nos connaissances langagières « proviennent directement d'une grammaire universelle qui fait partie de notre patrimoine génétique » [Tellier 1995, p. 27].

Ex : *Corpus de syntaxe*

- 1 - (?) Voici les livres que tu as rangés sans même avoir lus.
- 2 - (?) Quels sont les livres que tu as rangés sans même avoir lus ?
- 3 - * Tu ne devrais pas ranger ces livres sans même avoir lus.
- 4 - * Tu ne devrais pas les ranger sans même avoir lus ?
- 5 - Tu ne devrais pas ranger ces livres sans même les avoir lus.
- 6 - Tu ne devrais pas les ranger sans même les avoir lus ?

Les phrases 1 et 2 correspondent à des *constructions à vides parasites*¹, rarement employées, non mentionnées dans les grammaires traditionnelles et non enseignées dans les écoles françaises. Pourtant, tout locuteur du français a une intuition claire sur ces phrases, au sens où il est en mesure de trancher sur la grammaticalité des unes et l'agrammaticalité des autres, ce qui amène l'auteur à poser la question de savoir comment les locuteurs du français peuvent « apprendre » ce type de structure à leur insu ? Dans la lignée chomskyenne

¹ Ces constructions sont soumises à des contraintes syntaxiques spécifiques ; en particulier, le complément du premier verbe doit être relativisé ou questionné, ce qui explique l'agrammaticalité des deux phrases intermédiaires [Tellier 1999, 2002, p. 27].

[Chomsky 1957, 1965], elle répond que les locuteurs ont une connaissance de leur langue qui va bien au-delà de ce qui est accessible et généralisable à partir des seuls stimuli auxquels ils sont exposés. C'est cette connaissance implicite de la langue qui permet de conclure à l'existence d'une grammaire universelle innée.

Le *corpus électronique* correspondant à notre support d'analyse, nous consacrons la section suivante à sa définition.

1.2 Le corpus électronique

1.2.1 Le corpus informatisé

La définition de *corpus informatisé* la plus couramment citée est certainement celle de Sinclair ; ce qui s'explique probablement par l'intérêt précurseur de ce chercheur pour la linguistique de corpus (*cf.* Chapitre 1) et le caractère fondateur des travaux qu'il a publiés : « [a corpus is] a collection of pieces of language that are selected and ordered according to explicit linguistics in order to be used as sample of the language. » [Sinclair 1996, p. 4].

La définition proposée par Habert reprend largement celle de Sinclair, tout en introduisant une dimension plus linguistique. En effet, elle restreint les données possibles aux occurrences d'une langue particulière, sélectionnées selon une méthode mixte multidimensionnelle : « Un corpus est une collection de données langagières qui sont sélectionnées et organisées selon des critères linguistiques et extralinguistiques explicites pour servir d'échantillon d'emplois déterminés d'une langue. » [Habert 2000, p. 13].

Nous nous référons également à la définition de Rastier, qui, outre le fait de faire explicitement référence à l'aspect électronique du corpus, présente également l'intérêt de spécifier deux aspects épistémologiques de la linguistique de corpus : « Un corpus est un regroupement structuré de textes intégraux, documentés, éventuellement enrichis par des étiquetages, et rassemblés : (i) de manière théorique réflexive en tenant compte des discours et des genres, et (ii) de manière pratique en vue d'une gamme d'applications. » [Rastier 2005, p. 32].

Bien que de traditions linguistiques différentes (linguistique de corpus, TAL et sémantique), ces trois définitions s'accordent sur un même principe descriptif. Un corpus électronique se caractérise par une nature, une structure et une finalité.

Une nature : l'ensemble constitué par les documents recueillis est composé de données langagières écrites (romans, journaux, publicités, lettres...) et/ou orales (interviews, conversations, monologues...).

Une structure : cette notion convoque deux réalités, la sélection des données d'un côté, et de l'autre, la mise en conformité et l'enrichissement du corpus.

La sélection (organisation, structure) implique immédiatement la question du choix des données à assembler et pose la question de leur représentativité par rapport à l'objectif visé de la recherche : langue générale ou langue de spécialité. Les critères appliqués sont d'ordre linguistique (traits linguistiques : temps du passé, troisième personne... ; paramètres

situationnels : canal, format...) et extralinguistique (domaine : médecine, bourse... ; genre : article de recherche, journaux...).

La mise en conformité d'un corpus électronique correspond par exemple à l'harmonisation du codage de ses caractères (**Ex** : format UTF-8), et/ou à l'homogénéisation de ses informations par structuration normée (**Ex** : balisage DTD-TEI). L'enrichissement correspond par exemple à l'encodage des informations dans un langage de normalisation (**Ex** : XML), à l'annotation linguistique des parties du discours (**Ex** : nom, nom propre, verbe...), ou encore au marquage du lemme (**Ex** : cheval – chevaux, aller – va...). De fait, un corpus est « encodé de manière standardisée et homogène pour permettre des extractions non limitées à l'avance » car « la simple la simple existence sur support électronique ne fait pas d'un ensemble de textes un corpus électronique. Encore faut-il que ce document obéisse à des conventions de représentation, de codage répandues, voire consensuelles, qui permettent la transmission et la réutilisation des données textuelles en cause ». [Habert et al. 1997, p. 145].

Une finalité : d'une façon générale, il s'agit d'obtenir un échantillon représentatif soit d'une langue générale, soit d'une langue de spécialité. La linguistique de corpus est mieux établie dans le domaine de la langue générale que dans les langues de spécialités, beaucoup de ce qui s'écrit à propos du travail sur corpus² traite des corpus de langue générale, appelés *corpus de référence*. Pour la conception de travaux sur *corpus spécialisé*, il est souvent nécessaire de reprendre et d'adapter les procédures établies pour des études portant sur la langue générale.

1.2.2 Le corpus de référence

« Un corpus de référence est conçu pour fournir une information en profondeur sur une langue. Il vise à être suffisamment grand pour représenter toutes les variétés pertinentes de cette langue et son vocabulaire caractéristique. » [Sinclair 1996, p. 10].

Le corpus de référence est consacré à une langue naturelle (**Ex** : Brown corpus : anglais américain ; Lancaster-Oslo-Bergen Corpus : anglais britannique). Il tend à représenter la diversité des usages de la langue choisie. À ce titre, il est constitué d'un ensemble de données dont les conditions de production et de réception sont représentatives d'une grande variété de situations de communication (orale : monologue, interview ; écrite : lettre, roman...) et de types textuels (exposé scientifique, fiction narrative, reportage...). Il permet la constitution de sous-corpus en registre³ pour des analyses contrastives, par exemple [Péry-Woodley 1995].

² Dès lors, nous employons *corpus* pour *corpus informatisé*.

³ *Registre* est employé au sens de Biber [Biber 1995], pour une conception élargie des genres, initialement entendus comme des catégories de textes distingués spontanément par les locuteurs confirmés d'une langue. Par exemple, pour Biber [Biber 1988, 1989] les genres de l'anglais incluent les romans, les articles de journaux, les éditoriaux, les articles de recherche, etc. L'idée d'une possibilité de catégorisation intuitive des textes est communément admise : « la catégorisation des textes fait partie des activités cognitives spontanées des sujets. [...] Sans l'existence de catégories, notre appréhension des énoncés produits serait probablement impossible ». [Adam 1992, p. 6]. Pour autant, si le principe de catégorisation est intuitif, le résultat de ce processus l'est beaucoup moins ; ce qui se traduit par une grande instabilité terminologique entre les notions de registre et de genre : « les frontières entre registre et genre sont inévitablement floues et il serait vain de chercher à en établir une liste exhaustive ». [Péry-Woodley 2000, p. 160]. Nous utilisons désormais *registre*.

En outre, le corpus de langue générale est souvent *ouvert*, son contenu est sans cesse augmenté de nouvelles données, ce qui autorise à terme des analyses diachroniques (néologismes, emplois morphologiques privilégiés...).

Enfin, le corpus de langue générale est de grande taille ; il dépasse aujourd'hui plusieurs millions d'occurrences. Nous sommes entrés dans l'ère des *méga-corpus* ; nous présentons, à titre d'exemple, le plus récent d'entre eux.

Le *British National Corpus* – BNC – [Aston et Burnard 1998] est consacré à l'anglais britannique. Il a été créé par un consortium composé d'Oxford University Press, l'éditeur de dictionnaires Longman et Chambers, des centres de recherche des universités de Lancaster et d'Oxford et de la British Library. Il est conçu dès sa création, en 1991, comme un corpus multi-usage, à visées lexicographiques et éducatives. Concrètement, la représentativité de la diversité des usages est respectée, car :

- pour l'oral (10%), des conversations spontanées ont été recueillies à partir d'un échantillonnage démographique selon l'âge, le sexe, le groupe social et la région. Des transcriptions d'interactions orales dans divers domaines ont été intégrées. On répertorie le domaine des affaires (réunions, prises de parole syndicales, consultations médicales ou légales), de l'éducation et de l'information (cours et conférences, informations radiotélévisées), des prises de parole publiques (sermons, discours politiques, discours parlementaires et légaux), des loisirs (commentaires sportifs, réunions de clubs) ;
- en ce qui concerne l'écrit (90%), ont été pris en compte le domaine (textes informatifs à partir de 1975 et textes de fiction à partir de 1960), le support (livres, périodiques, discours), la datation et la diffusion (sélection parmi les listes des meilleures ventes, celle de prix littéraires, les indications de prêts en bibliothèque, etc.).

Techniquement, le BNC compte environ cent millions de mots. Il est fourni avec un logiciel d'exploration nommé SARA [Aston and Burnard 1998]. Il est balisé en SGML et annoté grammaticalement, pour en faciliter la consultation en termes de registres.

Cette possibilité de consultation et de catégorisation par registres nous amène à préciser que, pour bien des chercheurs, les corpus de référence s'apparentent plus à des *bases textuelles* qu'à des vrais corpus, car c'est le principe du choix raisonné parmi les composants disponibles qui crée un corpus [Habert 2000]. Par conséquent, les méga-corpus représentent davantage des réservoirs à corpus, dans la mesure où « les données signalétiques attachées à chaque composant permettent de réaliser à façon un corpus répondant à une recherche particulière » [Illouz et al. 1999, p. 44].

1.2.3 Le corpus spécialisé

La constitution d'un corpus spécialisé a pour origine une hypothèse de recherche : « Il ne peut y avoir de recherche sans que soit formulée une hypothèse de recherche qui va guider non seulement la constitution du corpus mais sa consultation. » [Martin 1997a, p. 80].

Il est généralement consacré à une situation de communication ou à un domaine de la connaissance [Habert et al. 1997] et l'hypothèse de recherche porte sur la langue de spécialité en usage dans le domaine, dont il doit être représentatif. Parce que le corpus spécialisé se fait

le miroir d'un domaine, il se doit d'en représenter la singularité. Les critères de sélection des données sont généralement fondés sur une *configuration domaine-genre* (**Ex** : articles de recherche en biochimie), pour reprendre la terminologie de Péry-Woodley [Péry-Woodley 2000 ; Péry-Woodley and Rebeyrolle 1998]. Le corpus spécialisé doit également être homogène, c'est-à-dire que les sources de données sont diversifiées [Atkins and al. 1992] ; ce qui importe n'est pas tant le nombre de mots que la quantité de documents rassemblés et leur variété. En outre, diversifier les sources laisse espérer une répartition plus équilibrée des mots dans le corpus. Les textes regroupés doivent rester entiers, ne pas être tronqués, pour ne pas perdre la cohérence du document et en garantir sa valeur conceptuelle.

Enfin, il n'y a pas de consensus sur la taille optimale attendue pour un corpus spécialisé, elle oscille le plus souvent entre cinq cent mille [Williams 1999b] et un million de mots [Pearson 1998].

Nous prenons pour exemple le corpus *Biveg*, consacré aux plantes parasites. Ce thème concerne deux communautés scientifiques : la biologie moléculaire et la physiologie végétale. Créé en 1996 par Geoffrey Williams pour les besoins d'un projet de recherche ponctuel, il a été conçu initialement pour étudier des articles dans une optique pédagogique. Par la suite, le corpus de base a été augmenté afin de mesurer les apports respectifs des deux langages de domaine (la biologie et la physiologie) en contexte pluridisciplinaire, dans le but de créer un dictionnaire pédagogique correspondant à ce domaine discursif.

Concrètement, la représentativité (de la singularité) du thème est respectée, puisque le corpus est constitué d'articles de recherche publiés soit dans les revues internationales de la communauté, soit dans les actes du Congrès du Cordoue sur les plantes parasites [Moreno et al. 1996]. Techniquement, *Biveg* est un corpus ouvert, actuellement composé de 154 articles de recherche ; il compte environ 435 000 mots. Il est balisé, étiqueté grammaticalement et lemmatisé.

Ainsi, on retient pour le corpus spécialisé tout regroupement de données langagières créé à des fins spécifiques et représentatif d'une situation de communication ou d'un domaine de pratique. Dans la perspective sémantico-discursive de cette recherche, nous retenons la procédure qui implique :

- que le domaine des textes inclus dans le corpus soit défini et délimité ;
- que les textes soient représentatifs de ce domaine pour garantir l'authenticité des conclusions qu'on en tire.

Or « il n'est pas possible de construire un corpus correct si nous n'avons pas défini clairement la réalité que nous cherchons à décrire » [Williams 1999b, p. 38]. La « réalité » dont il est question ici convoque les concepts de « communauté » et de « domaine » évoqués auparavant, dont le besoin de circonscription devient évident. En effet, c'est en délimitant une communauté, un domaine que les usages linguistiques en vigueur peuvent être explicités, ce qui implique préalablement de trouver un modèle théorique capable de guider cette démarche.

2 Circonscrire un domaine

2.1 Langue spécialisée vs langue de spécialité⁴

Nous proposons dans cette sous-section d'introduire modestement la notion de *langue de spécialité* et d'évoquer très brièvement les problématiques communément associées, au regard de son statut par rapport à la langue générale. Trouver une définition tranchée du concept semble utopique, car nombreuses sont les hésitations qui en discréditent l'existence : « Le caractère extralinguistique (externe) des spécialités, aussi bien que les difficultés de saisir des corrélations majeures entre le linguistique et l'extralinguistique, mènent au doute relatif à la définissabilité de la langue de spécialité, à l'acceptabilité d'un tel concept. » [Kocourek 1991, p. 19]. Malgré ces réserves, trois notions apparaissent comme constitutives de la définition de ce concept [Kocourek 1991 ; Lerat 1995] : une origine, une nature et une fonction.

Une origine : la langue de spécialité appartient à un domaine, à un champ d'expérience particulier.

Une nature : la langue de spécialité figure une « variété » de la langue générale ; on lui attribue toutes les caractéristiques de la langue générale (terminologiques, syntaxiques...), mais restreintes par la spécialité.

Une fonction : la langue de spécialité est employée pour communiquer ; elle a une vocation informationnelle et sert en ce sens à véhiculer des connaissances.

Les problématiques liées à l'aspect linguistique de la langue de spécialité s'intéressent au caractère interne à la langue (lexique et syntaxe) et font de la langue de spécialité un objet langagier distinguable de la langue générale. Les trois notions constitutives de la définition de langue de spécialité permettent de caractériser ce concept au moyen des éléments de base suivants : spécialité, connaissances/communication et sous-langage. Si les deux premiers points semblent tout à fait explicites, le dernier reste plus ou moins ambigu. En effet, la langue de spécialité est « une variété de la langue générale », on lui attribue le statut de sous-langage [Müller-Thurau 1985, p. 172], mais les caractéristiques que recoupe ce statut sont loin d'être évidentes. On trouve dans la littérature une idée élégante attribuée à Coseriu [Coseriu 2000] selon laquelle, par exemple, le langage de la science, le langage technique est simplement une possibilité du langage en général. Au premier abord, l'idée séduit mais ne résout pas la question du statut accordé à la langue de spécialité par rapport à la langue générale et, plus précisément, la nature des relations qu'entretiennent le lexique et la grammaire de la langue de spécialité par rapport à la langue générale.

Du point de vue lexical, chaque langue naturelle opère un découpage particulier du réel dont témoigne son vocabulaire. Par ailleurs, une communauté bâtit une représentation du monde supposée correspondre à la réalité des préoccupations de celle-ci, du moins telle que la définit

⁴ Le concept de langue spécialisée réfère préférentiellement à l'acception de sous-langage, tandis que celui de langue de spécialité réfère davantage à la langue en usage dans un domaine de pratique [Habert 2000]. Aussi, les notions de langue spécialisée et de langue de spécialité sont à la fois communes et particulières aux deux théories présentées ensuite. Toutefois, nous utilisons dès à présent le terme de langue de spécialité, eu égard au modèle théorique des communautés de discours retenu pour cette étude, sauf pour la section consacrée à la théorie des sous-langages, où nous conservons le terme de langue spécialisée.

la collectivité à un moment donné, et l'une des manifestations linguistiques de cette activité est la constitution de terminologie, ou ensemble plus ou moins structuré de termes spécialisés. Or « la terminologie a ceci de particulier qu'elle appartient aussi au lexique de la langue naturelle » [Mortureux 1985, p. 827]. Il n'y a donc pas de cloisons étanches entre le lexique de la langue naturelle et le lexique spécialisé. L'interférence entre les deux est indiscutable, ce qui rend impossible la délimitation d'une langue de spécialité par l'entrée lexicale.

D'un point de vue grammatical, la problématique est la même, à ceci près qu'il est légitime de se demander auparavant si la grammaire des langues de spécialité diffère réellement des structures grammaticales communes à la langue générale. La réponse semble plutôt négative : « Les recherches dirigées par M. Phal ont montré qu'en matière de langue scientifique et technique une langue de spécialité comporte, à côté d'un lexique spécifique correspondant à une spécialisation très poussée, une matière linguistique, tant lexicale que grammaticale. » [Lerat 1995, p. 98]. Un tel constat autoriserait à reconnaître qu'il existe deux instances grammaticales : une grammaire commune et une grammaire spécialisée. Il est généralement admis que la grammaire du français spécialisé est essentiellement sélective ; elle est caractérisée par des schémas syntaxiques préférentiels ou contraignants. La spécificité des grammaires spécialisées réside dans la manière dont elles mobilisent et actualisent les ressources de la langue naturelle et surtout dans la combinatoire particulière qui se manifeste au niveau de l'unité transphrastique. Aussi, comme pour le lexique spécialisé, des interférences existent entre la grammaire spécialisée et la grammaire générale, ce qui rend également difficile la délimitation d'une langue de spécialité par l'entrée grammaticale.

Dans cette perspective et selon les deux critères, lexical et grammatical, dont il vient d'être question, soit le lexique et la grammaire de la langue de spécialité sont inclus dans la langue générale, soit les deux ensembles se croisent. En aucun cas la langue de spécialité ne peut décemment constituer un module tout à fait indépendant.

Les difficultés évoquées quant aux critères opérationnels de délimitation de la langue de spécialité témoignent de la nécessité de définir le concept sur la base de critères externes ; ce qui ramène la difficulté à la nécessité de circonscrire un domaine de pratique. Pour constituer un corpus spécialisé, il s'avère donc nécessaire de pouvoir définir la communauté dans laquelle la langue de spécialité est usitée. Deux cadres théoriques permettent un tel exercice : la théorie des sous-langages et la théorie des communautés de discours.

2.2 Théorie des sous-langages

La théorie des sous-langages – SL – est inspirée d'un modèle grammatical générativiste. Le terme lui-même est mentionné pour la première fois dans les années 1960 par Harris concernant la grammaire des langues spécialisées : « Certains sous-ensembles propres des phrases d'un langage peuvent être fermés pour certaines (ou toutes les) opérations définies dans le langage, et constituer un sous-langage de ce langage. » [Harris 1968, cité dans Williams 1999b, p. 44]. À l'époque, le défi consiste notamment à expliquer pourquoi certaines structures grammaticales de l'écrit scientifique ne se laissent pas décrire par les règles transformationnelles du langage général, principe théoriquement universel. Les SL étant des « sous-ensembles fermés », des modules possiblement indépendants du langage général, leur lexique est supposé fini, c'est-à-dire théoriquement dénombrable : « There is no upper bound

on the size of vocabulary of a natural language. In contrast to it the vocabulary of a sublanguage is theoretically finite. » [Moskovich 1982, p. 193]. Un SL correspondant à une « quantité finie », la TSL considère qu'il est possible de délimiter un SL. Cette idée, fondamentale dans ce modèle, suppose de savoir sur quels types critères repose la délimitation. La réponse apportée par les auteurs reste floue, l'origine des SL et les critères de temps n'étant pas pris en considération. Toutefois, quelques éléments de réponse se dégagent autour de critères linguistiques et/ou extralinguistiques.

Sur le plan linguistique, on retrouve les problématiques posées par le lexique et la grammaire (cf. Section 2.1). Pour la TSL, autant la quantification des termes du lexique d'un SL est supposée possible, autant ses structures grammaticales sont supposées en partie identiques à celles de la grammaire du langage général : « [...] Tandis que les phrases des langages-objets scientifiques sont incluses dans le langage total, la grammaire de ces sous-langages a une intersection non vide (plutôt que est incluse) dans la grammaire du langage total. » [Harris 1968, cité dans Williams 1999b, p. 44]. Le point de vue lexical pose problème dans la mesure où il se soustrait au principe dynamique des langues vivantes, selon lequel elles intègrent perpétuellement de nouveaux mots (**Ex** : par emprunt ou par dérivation morphologique). Tandis que le point de vue grammatical pose la question de la prédictibilité des structures propres au SL, une fois inventoriées : « Les hypothèses de Harris (1988) sur la lexicogrammaire des articles de recherche ne sont pas reproductibles sur un corpus car la réalité est plus expressive que la grammaire simplifiée de Harris. » [Williams 1999b, p. 50].

Pour surpasser ces difficultés, Kittredge introduit la notion de *conventionnalisation* de l'usage des termes, qui concerne simultanément l'aspect lexical et grammatical : « [...] A new term or grammatical construction does not become a true part of the sublanguage until its use has been conventionnalized by the community of speakers. » [Kittredge 1982, p. 110]. L'idée séduit mais se révèle à nouveau tendancieuse, par la question de l'instabilité perpétuelle des éléments d'un SL qui mettent en doute la possibilité d'une réelle délimitation. De fait, il s'agit du principal écueil auquel se heurte la théorie classique des SL. Aussi, les études plus récentes semblent moins restrictives sur cette question :

« What is sublanguage ?

Part of a natural language

Specialized form

Behaves like a complete language

Used in special circumstances – expert communication

Limited to particular subject domains

Sublanguage are not artificial languages

fuzziness of boundaries of sublanguages

fuzziness of boundaries of natural languages... » [McNaught 1992, p. 3]

Pour McNaught, il existe une interactivité entre les SL et le langage général, mais également des SL entre eux. Par exemple, la biologie et la chimie diffèrent par les structures syntaxiques des termes et par la terminologie nécessaire à la description des phénomènes particuliers aux domaines, mais la biochimie possède des tournures syntaxiques et terminologiques communes aux deux disciplines.

Sur le plan extralinguistique, la TSL envisage deux types de critères externes à la langue spécialisée elle-même : l'un fondé sur une sous-catégorisation des spécialités par thème, l'autre basé sur l'usage des locuteurs. La sous-catégorisation progressive des spécialités par thème reprend exactement la procédure de classement Dewey des documents dans une

bibliothèque. Cette classification décimale répartit l'ensemble du savoir humain en dix grandes classes, numérotées de 000 (généralités) à 900 (histoire et géographie). Chaque classe est séparée en dix divisions (**Ex** : 000-009 Savoir et communication, 010-019 Bibliographies, 020-029 Bibliothéconomie et science de l'information, 030-039 Encyclopédies générales...) et celles-ci de nouveau en dix subdivisions. Un indice de classification comporte trois chiffres. Si un indice dépasse trois chiffres, un point sépare les trois premiers des suivants. À l'intérieur des grandes classes du système Dewey, les sujets sont répartis suivant un ordre qui va du général au particulier. C'est donc au début de chaque centaine et de chaque dizaine qu'on retrouve les manuels, les traités généraux et les encyclopédies. Les sujets très spécialisés viennent par la suite et se voient attribuer un indice plus développé. Aussi, pour accéder à un SL, il faut extraire les textes d'une source uniforme, c'est-à-dire d'une sous-classe. Toutefois, la question des critères de choix reste entière. Le fait de rassembler des textes issus de la chirurgie ne permet pas de savoir lesquels prendre.

Le critère externe de l'usage des locuteurs repose sur des « habitudes partagées » entre locuteurs appartenant à un même domaine : il existerait des « shared habits of words usage on part of the speakers » [Kittredge 1982, p. 110]. Mais, là encore, ce critère ne définit guère les SL de façon exhaustive. En outre, la notion d'appartenance à un domaine n'est pas un concept défini dans la théorie classique des SL.

Ainsi, sur le plan des critères internes à la langue, bien que dans ses versions les plus récentes la TSL tente d'intégrer les fonctionnements de la langue naturelle, les SL restent difficiles à circonscrire. La théorie ne parvient jamais à établir ni à justifier la relation qui unit le SL au langage général. Sur le plan des critères externes, la théorie donne l'impression que les SL existent de fait, sans qu'il y ait de difficulté à les révéler. La TSL préconise en effet de délimiter une langue spécialisée et d'accéder ainsi à un champ de pratique. Or cette démarche se révèle trop restrictive et cloisonnée pour être exploitable dans l'optique d'une constitution de corpus spécialisé. Nous avons besoin d'un modèle plus descriptif, capable de proposer des critères tangibles.

2.3 Théorie des communautés de discours

La théorie des communautés de discours – TCD –, contrairement à la TSL, cherche non pas à accéder d'emblée à une langue spécialisée mais à circonscrire une communauté scientifique [Tognini-Bonelli 2001]. Williams justifie ainsi cette démarche : « Dans la mesure où les variétés différentes d'une langue sont utilisées dans des situations sociales différentes et où le locuteur natif peut passer rapidement d'une variété à une autre, il sera difficile de définir une norme pour laquelle la valeur de la langue de référence n'est autre qu'une valeur sociale. » [Williams 1999b, p. 50].

Une communauté de discours – CD – se définit comme un *groupement sociorhétorique* [Swales 1990, pp. 24-27], ce qui a pour conséquence directe que le choix des critères de délimitation s'applique non pas à la langue de spécialité en usage dans la communauté mais à la communauté elle-même.

Six critères permettent de postuler l'existence d'une CD et de la circonscrire :

- 1- un but commun accepté par tous les membres de la communauté ;
- 2- des mécanismes d'interaction entre les membres ;

- 3- des mécanismes de participation garantissant l'information et le retour d'informations ;
- 4- l'utilisation et la possession d'un ou de plusieurs genres ;
- 5- l'acquisition d'un lexique spécifique ;
- 6- un seuil de membres.

On peut répartir ces six critères de délimitation autour des deux composantes, sociologique et rhétorique, mobilisées par le concept *sociorhétorique*.

Du point de vue sociologique (critères 1, 2, 3, 6), l'existence d'une CD repose sur un accord mutuel des membres de la communauté au sujet de concepts appliqués à un objet d'intérêt commun. Si l'objet fait appel à des concepts partagés à l'intérieur d'une discipline, il s'agit d'une communauté de domaine [Williams 2002]. Dans ce cas, bien souvent, le corpus est composé des textes publiés dans la revue de la communauté. La constitution du corpus revient donc à choisir la communauté à étudier et à déterminer le nombre d'années à inclure dans le corpus. À l'inverse, si l'objet d'étude est pluridisciplinaire, il s'agit alors d'une communauté de thème. Dans ce cas, son caractère transdisciplinaire implique plusieurs domaines en présence. Dès lors, la priorité en matière d'assemblage est d'équilibrer les sources et de diversifier les auteurs. La représentativité du corpus est assurée si plusieurs publications par spécialité sont intégrées [Williams 1999b]. Les mécanismes d'interactions et de participations entre les membres ne sont pas toujours faciles à décrypter. Ils varient d'une communauté à l'autre et sont par exemple dépendants des procédures de réunion entre les membres, des moyens de télécommunication mis à leur disposition, de leurs habitudes de correspondances (papier, électroniques)...

Du point de vue rhétorique (critères 4, 5), une communauté échange au travers de genres préférés (articles de recherche, résumé...) ; Swales caractérise ainsi la notion de genre : « A genre comprises a class of communicative events, the members of which share some set of communicative purposes. These purposes are recognized by the expert members of the parent discourse community, and thereby constitute the rationale for the genre. » [Swales 1990, p. 58]⁵. Les moyens du discours sont des signes d'appartenance à une CD, aussi, pour intégrer une communauté il convient d'adopter les règles de discours qui y sont en vigueur. Une CD possède des termes et des structures syntaxiques propres au domaine, elle attribue des significations spécifiques à certains termes et privilégie des phrases types. Toutefois, la communauté de discours n'est pas définie par ce lexique et ces choix syntaxiques, contrairement aux sous-langages [Harris 1968].

En effet, la TCD rejette l'idée du statut semi-autonome de la langue de spécialité, dérivé de la langue générale, et considère qu'il s'agit seulement de préférences lexico-grammaticales dans un contexte sociorhétorique professionnel, une forme de « sociolecte professionnel », entendu comme les variations dénotant une appartenance sociale.

Ainsi, « la communauté de discours est un ensemble ni fermé, ni fini, mais définissable », elle « a une existence réelle, reconnue par ses membres et peut par conséquent être décrite de

⁵ À noter que la définition de *genre* proposée par Swales [Swales 1990] correspond à celle de *registre* proposée par Biber [Biber 1995].

l'intérieur, au lieu de faire appel à une classification imposée et d'essayer de prouver le caractère réel de cette catégorisation externe » [Williams 1999, p. 52]. En outre, la TCD privilégie l'aspect interdisciplinaire des communautés, puisqu'elle tient compte des influences mutuelles des CD entre elles, de la contamination et de l'interférence réciproque des domaines en présence. Par exemple, la TCD parvient à délimiter un domaine tel que la biochimie, intersection entre la biologie et la chimie.

2.4 Justification du choix théorique

Les contraintes de constitution d'un corpus électronique spécialisé sont principalement fonction de l'hypothèse de recherche qui en a motivé le besoin et par répercussion du paradigme théorique dans lequel elle s'inscrit. La création du corpus TAL(N) est née de « l'incompatibilité apparente » de deux préceptes (*cf.* Chapitre 4) : le premier en lien avec l'objectivité du discours scientifique spécialisé et le second relatif au caractère évaluatif de la langue générale, principe théorique fondateur en sémantique des possibles argumentatifs – SPA – (*cf.* Chapitre 2). L'ancrage de notre recherche dans ce modèle sémantico-discursif nous a obligés à opter pour un processus de construction du corpus électronique spécialisé fondé sur des critères externes, visant à délimiter le domaine du TAL et non à accéder d'emblée à une langue de spécialité, afin d'éviter l'écueil de la circularité des contraintes dans la constitution du corpus. D'une façon générale, Pearson [Pearson 1998] et Martin [Martin 1997a] nous mettent en garde contre le fait de construire un corpus uniquement pour tester l'hypothèse de travail, car le risque est grand de trouver ce qui est cherché et d'ignorer le reste comme bruit. Une telle démarche induit un biais dans la consultation et l'analyse du corpus, biais propre à générer des conclusions abusives. En particulier, l'analyse sémantico-discursive exclut :

- d'une part de constituer un corpus fondé (même en partie) sur des facteurs linguistiques internes [Biber and Finegan 1986 ; Biber 1988, 1993] ou sur des patrons lexico-syntaxiques à la manière des sous-langages [Harris 1968] ; la démarche serait circulaire et mènerait à des conclusions faussées ;
- d'autre part de constituer un corpus basé uniquement sur des critères externes situationnels, en regroupant les textes selon leur registre. Cette démarche est fondée quand elle s'intègre à une approche multidimensionnelle, pour l'émergence automatique de type de texte, par exemple [Biber 1988, 1993, 1995 ; Biber et al. 1994, 1998], mais trop approximative employée seule [Habert et al. 1997], dans la mesure où elle ne désigne pas objectivement les sources de données textuelles à rassembler.

Dans cette perspective, la TCD est adaptée à nos besoins, puisqu'elle permet de circonscrire une communauté de l'intérieur sans recourir uniquement à des critères internes de sélection linguistique et sans adopter une démarche de sélection externe des textes, basée uniquement sur une catégorisation par registre, même si celle-ci est inévitable [Rastier 2005].

Par ailleurs, la TCD représente une contribution empirique appréciable, puisqu'elle rassemble des critères de sélection stables pour circonscrire un domaine tout en désignant les sources de données textuelles à inclure dans le corpus. En outre, elle permet de mieux maîtriser les critères externes de sélection textuelle restant à la charge du chercheur, même s'ils demeurent partiellement subjectifs.

Ainsi, la construction d'un corpus spécialisé s'accompagne de contraintes particulières, de même que la démarche qui consiste à définir une CD. Il semble donc nécessaire de hiérarchiser les différentes contraintes de constitution inhérentes à ces deux paramètres, selon un principe chronologique d'exécution. Outre la nécessité de formuler une hypothèse de recherche et de définir un but spécifique (*cf.* Chapitre 4), il s'agit :

- de choisir un domaine de pratique ;
- de soumettre le domaine de pratique visé à l'hypothèse des CD, ce qui revient à délimiter la communauté visée en fonction des six critères posés par la TCD ;
- de déterminer s'il s'agit d'une communauté de thème ou de domaine ;
- d'établir les critères de sélection des textes entiers (écrits ou oraux), permettant une collecte des textes qui soit représentative du domaine.

3 Constitution du corpus TAL(N)

3.1 Délimitation du domaine scientifique TAL

3.1.1 Le domaine pressenti

Nous avons orienté notre choix vers le domaine interdisciplinaire du TAL, à l'interface entre l'informatique et la linguistique. Ce domaine de pratique scientifique est dynamique en France ; une association loi 1901, l'ATALA (Association pour le traitement automatique du langage)⁶ lui est adjointe. Les communications scientifiques du TAL sont essentiellement écrites, ce qui satisfait nos perspectives d'analyses textuelles, et publiées en français, langue maternelle sur laquelle nous travaillons. Elles sont nombreuses, ce qui offre une double perspective scientifique confortable. D'une part, le regroupement d'une quantité de données suffisante est possible et, d'autre part, cela laisse espérer une certaine diversité des auteurs permettant de décupler les possibles subjectivités et d'accéder de surcroît à un large panel lexical.

Par ailleurs, l'espace géographique dans lequel se déroule cette recherche (laboratoire informatique de Nantes Atlantique – LINA) compte plusieurs membres de la communauté TAL, ce qui nous permet de solliciter au besoin la compétence linguistique de ces locuteurs et de bénéficier de leur expertise sur le domaine.

3.1.2 La communauté de discours TAL

La vérification empirique proposée ci-dessous s'attache à valider l'existence d'une CD TAL. L'exercice consiste à confronter systématiquement la communauté scientifique pressentie aux six critères définis dans la TCD.

⁶ <http://www.atala.org/>

3.1.2.1 Un but commun

L'article 2 des statuts de l'ATALA atteste d'une orientation scientifique unique partagée par tous les membres de l'association.

« L'association a pour but de favoriser l'étude des problèmes théoriques et pratiques du traitement automatique des langues et de faciliter les échanges scientifiques dans ce domaine. Pour atteindre ces buts elle se propose en particulier :

- 1- d'assurer l'information et la diffusion des connaissances sur le traitement automatique des langues et sur les domaines connexes, notamment par la publication d'une revue ou d'un bulletin, d'ouvrages et de tous autres documents intéressant son domaine ;
- 2- d'organiser des rencontres dans le domaine du traitement automatique des langues (journées d'études, colloques, écoles d'été, etc.) ;
- 3- d'établir des contacts avec tous autres groupements s'intéressant aux mêmes sujets qu'elle-même ;
- 4- de contribuer par tous moyens appropriés à la production et à l'évaluation des connaissances dans le domaine du traitement automatique des langues. »

3.1.2.2 Des mécanismes d'interaction

Concernant les mécanismes d'interaction entre les membres, il existe des interactions administratives en lien avec l'existence de l'association. Il s'agit des assemblées générales, des conseils d'administration (CA), des échanges au sein du bureau et des organes scientifiques et techniques créés par le CA.

Il existe parallèlement des interactions scientifiques, en lien avec les organes scientifiques et techniques créés par le CA. Il s'agit de la revue *TAL*, des colloques (TALN, RECITAL, des conférences associées) et des journées d'études, qui offrent également un cadre officiel d'échanges pour les membres du domaine.

Nous pouvons aussi mentionner les échanges plus informels rendus possibles par l'utilisation du courrier électronique dans le cadre de la liste de diffusion LN et du forum LN-forum.

3.1.2.3 Des mécanismes de participation

Le site de l'ATALA est un lieu d'information sur le domaine du TAL. Le lien [Annuaire](#) permet par exemple d'accéder à des rubriques aussi diversifiées que les formations universitaires, les équipes de recherches, les thèses (soutenues ou annoncées), les outils pour le TAL, les associations en TAL (internationale, allemande, espagnole) et les conférences en TAL.

La liste de diffusion et le forum permettent de garantir le retour d'information.

3.1.2.4 L'utilisation et la possession d'un ou de plusieurs genres

Il s'agit d'une communication entre spécialistes ou futurs spécialistes du domaine (professionnels, chercheurs, professeurs, maîtres de conférence, doctorants...) recourant à plusieurs genres (l'article de recherche, le compte rendu de lecture et le résumé de thèse...).

3.1.2.5 L'acquisition d'un lexique spécifique

Ce critère implique une double interprétation, selon que l'on considère l'*acquisition* comme un fait résolu ou comme un processus en marche.

Dans le premier cas, deux questions se posent :

- S'agit-il de montrer qu'il existe des termes de la communauté communs au langage naturel mais dont la signification s'est radicalisée dans le TAL ?

Ex⁷ : *contrôle* : le contrôle en intelligence artificielle traite de la conduite et de la supervision du raisonnement utilisé pour résoudre un problème donné et peut être résumé ainsi : à chaque instant du processus de résolution d'un problème donné, quelle action le système doit-il exécuter ?

Ex : *système de question-réponse* : c'est un système permettant d'extraire les parties d'un document les plus pertinentes par rapport à une question posée. Il s'articule autour d'un moteur de recherche enrichi en amont par un module d'analyse des questions et en aval par des modules robustes d'analyse des documents assurant l'extraction de passages réponses en mêlant l'exploitation de traits linguistiques (morphologique, syntaxique ou sémantique) et l'utilisation de critères numériques [Voorhees and Harman 1999].

- Ou bien s'agit-il de montrer qu'il existe des termes spécifiques à cette communauté, qui n'existent que dans le TAL et qui lui sont par conséquent spécifiques ?

Ex : *linguiciel* : logiciel traitant la langue naturelle sur les différents niveaux, morphologique, syntaxique et sémantique, en utilisant des bases de connaissances linguistiques.

Ex : *entité nommée* : « Ce terme regroupe les noms propres communément reconnus comme tels (la classe ENAMEX des conférences MUC), ainsi qu'un certain nombre d'entités qui ne sont pas toujours considérées comme des noms propres : les noms collectifs (les Français, les néandertaliens...), les maladies ou encore les noms de personnages mythiques ou fictifs (Hercule, Colombo...). » [Fourour 2001, p. 442].

Toutefois, on constate combien s'en tenir à cette interprétation semble réducteur et trahit peut-être l'esprit de la TCD. En effet, envisager l'identification d'un lexique spécifique comme étant un critère de caractérisation sinon de délimitation d'une communauté revient à accéder à la communauté par son lexique, même si dans le cas de la TCD cet aspect ne représente qu'un critère parmi cinq autres. Pour sortir de cette logique, nous proposons une interprétation supplémentaire de ce critère en termes de *processus d'acquisition lexical*. Dans ce cas, par *processus*, on écarte définitivement la question de la quantification des termes, pour laisser place à une logique évolutive du lexique spécialisé. Cette interprétation nous

⁷ Tous les exemples mentionnés dans cette partie ont été glanés au détour des lectures, soumis et validés ensuite par des membres de la communauté.

semble cohérente avec le concept même de communauté de discours ensemble *non fermé ni fini*. Mais comment prouver l'existence d'un processus d'acquisition lexical en cours ? Il nous semble intéressant de distinguer au moins deux cas :

- premièrement, le cas où la communauté s'approprie un concept scientifique et le terme apparenté, tout en lui spécifiant une signification ;

Ex : compositionnalité : « La question de la compositionnalité sémantique est en effet une question traditionnelle en linguistique et on peut se demander si elle présente un intérêt particulier pour le traitement automatique des langues. » « [La] compositionnalité définit une méthode de calcul du sens fort précieuse pour la compréhension automatique des langues dites naturelles. » [Nazarenko 1998].

- deuxièmement, le cas où la communauté crée un nouveau concept et le terme correspondant.

Ex : multimodalité⁸ : « Une fois reconnue la difficulté de trouver une définition unique, [...] en essence [...] la multimodalité est l'utilisation de deux ou plusieurs parmi les cinq sens en vue de l'échange d'informations. » Remarquons que cette définition semble mettre en porte-à-faux la notion de *multimédia*, avec laquelle elle se superpose. Concept à visée théorique servant à élaborer des systèmes multimodaux d'interaction homme-machine (**Ex** : CHAMELEON, YTTM).

3.1.2.6 Un seuil de membres

On peut concevoir deux seuils de membres, l'un fixé par le nombre d'adhérents à l'association, c'est-à-dire environ trois cents personnes depuis trois ans, l'autre, non dénombrable mais existant, fixé simplement par le nombre de laboratoires et d'équipes de recherche s'intéressant à ces questions et, par répercussion, incluant toutes les personnes physiques y travaillant ou participant aux réunions scientifiques.

Dans la mesure où les six conditions sont simultanément remplies, l'hypothèse est validée, le domaine du TAL constitue une CD et devient de ce fait une réalité descriptible.

3.1.3 Une communauté de domaine

La communauté TAL est une communauté de domaine constituée de plusieurs thèmes. En témoigne l'organisation thématique de la revue *TAL* elle-même, adressée aux spécialistes du domaine. Elle assure une fonction de cohésion. Dans les revues les plus récentes, chaque parution est dédiée à un sujet particulier dont les articles sélectionnés rendent compte.

Ex : « Dialogue » (revue *TAL*, Vol. 43(2), 2002) ; « Problèmes épistémologiques » (revue *TAL*, Vol. 43(3), 2002) ; « Modélisation probabiliste du langage naturel » (revue *TAL*, Vol. 44(1), 2003) et comporte régulièrement un article de revue hors-thème (classé « Varia »).

⁸ *Multimodality in Language and Speech Systems*, édité par Björn Granström, David House et Inger Karlsson, collection « Text, Speech and language Technology », Vol. 19, Kluwer Academic Publishers, Dordrecht, juin 2002.

Il arrive même que le thème de la revue puisse à nouveau se décomposer en plusieurs sous-thèmes.

Ex : « Linguistique de corpus » (revue *TAL*, Vol. 42(2), 2001) ; trois thèmes : « Annotations et analyse linguistique », « Textes thèmes et genres », « Corpus et ELAO » (enseignement des langues assisté par ordinateur).

3.2 Sélection des données textuelles

En validant l'hypothèse d'une CD, nous avons regroupé un certain nombre d'informations sur la communauté TAL propres à désigner plusieurs sources potentielles de données textuelles (*cf.* Annexe 1) : la revue *TAL* (les articles, les notes de lectures, les comptes rendus, les résumés de thèses), les actes des colloques (TALN, RECITAL, conférences associées) et les journées d'études thématiques organisées cinq fois par an. Ces multiples *formats textuels* génèrent presque autant de genres discursifs différents : l'article de recherche (de conférence), la note de lectures, le compte rendu, le résumé...

Consécutivement à l'application de la TCD, pour être représentatif de la communauté TAL, le corpus est constitué de textes appartenant aux différents supports référencés précédemment, mais lesquels choisir... *Not chance but choice*, peut-on lire dans le titre de l'article d'Engwall [Engwall 1994]... *Choisir* implique nécessairement dans la sélection des textes une forme de subjectivité que le chercheur doit combattre. De déductions en éliminations, les articles scientifiques de la revue *TAL* et de la conférence TALN se sont imposés.

Le corpus ne pouvait être composé des notes de lectures, des comptes rendus, des résumés de thèse et des actes de la conférence RECITAL, puisqu'ils sont principalement rédigés par les doctorants et jeunes chercheurs. Recueillir des textes issus de ces différentes sources aurait entravé la représentativité du corpus en surimposant les pratiques discursives des futurs membres du domaine. Nous avons également retiré les textes relatifs aux conférences associées et aux journées d'études pour leur caractère thématique. Leur présence aurait pu privilégier certains axes de recherches et certains auteurs, au détriment d'une représentativité globale de la communauté. En outre, collecter les données relatives aux journées d'études s'est vite révélé improbable, puisque les exposés et démonstrations ne font pas systématiquement l'objet d'un support électronique.

En revanche, les articles de la revue *TAL* et de la conférence TALN couvrent manifestement la communauté au sens le plus large. Premièrement, les auteurs sont variés (256 auteurs différents tous articles confondus, universitaires, chercheurs, professionnels, jeunes chercheurs, doctorants), ce qui garantit une bonne représentativité lexicale et une subjectivité potentielle tangible. Deuxièmement, nous évitons l'écueil qui consiste à privilégier un thème par rapport aux autres, *TAL* et *TALN* ayant pour but de présenter les avancées de la recherche TAL dans sa diversité.

4 Construction du corpus TAL(N)

4.1 Méthodologie d'obtention du corpus

4.1.1 Les fichiers

Les articles de la revue *TAL* et de la conférence TALN appartiennent à l'ATALA, organisme qui nous a donné l'autorisation de créer un corpus sur le domaine du TAL. Toutefois, l'ATALA n'ayant pas ces documents en sa possession, les fichiers ont été collectés auprès de chaque auteur ou groupe d'auteurs concernés, et l'exploitation légale des documents sous la forme numérisée a été possible sur acceptation de leur part.

À noter que pour la revue *TAL* 42.1 – Synthèse de la parole à partir du texte –, parue en 2001, nous avons obtenu pour neuf des onze fichiers concernés les dernières versions avant passage chez l'éditeur. Concrètement, il existe quelques infimes différences entre les fichiers utilisés dans le corpus TAL(N) et les articles réellement publiés dans la revue papier. Les variations sont d'ordre lexical (**Ex** : *un système de synthèse* – *un système **complet** de synthèse*), bibliographique (**Ex** : *procédures de calcul [?]* – *procédures de calcul [DUT 97]*), graphique (**Ex** : *comprenaient 2 procédures* – *comprenaient **deux** procédures*) et diacritique (**Ex** : *la prosodie (accentuation et intonation)* – *la prosodie : accentuation et intonation*).

Au fur et à mesure de leur réception, les fichiers ont été triés et stockés, c'est-à-dire répartis dans un répertoire reprenant leur type – *article de revue* ou *article de conférence* –, le numéro de la revue et l'année de publication.

Une fois tous les fichiers obtenus, ils ont été renommés pour en faciliter la manipulation. Nous avons uniformisé les intitulés. Les articles de revue ont reçu une étiquette du type AR_0001 et les articles de conférence, une étiquette du type AC_0001, la provenance des articles étant une information que nous ne voulions pas perdre en vue d'une étude contrastive possible. Le corpus TAL(N) est donc composé de deux sous-corpus. Ces informations sont répertoriées dans un fichier Excel, pour garantir la traçabilité des textes. Ce fichier Excel recense neuf types d'informations (*cf.* Annexe 2). Dans l'ordre d'apparition :

- le nom du fichier original – tel qu'il nous a été délivré ;
- le nouveau nom – celui que nous lui avons attribué ;
- la nature du texte (introduction ou article) – pour les articles de revue seulement ;
- l'auteur ou les auteurs – dans l'ordre figurant sur la publication papier ;
- le titre du texte ;
- le titre de la revue ;
- l'année de parution ;
- le numéro de la revue ;
- la partie thématique de la revue à laquelle appartient l'article.

Les fichiers ont généralement été réceptionnés au format .zip. Aucun d'entre eux n'a été endommagé par l'opération de décompression. Une fois décompressée, les fichiers étaient majoritairement sous format .ps ou .pdf. Leur conversion au format .txt a été effectuée *via* des scripts disponibles dans le domaine public (successivement appliqués) du type pstotext, pdftotext, rtftotext..., ce qui n'a pas toujours été simple ! Une fois au format .txt, les marques de mise en page ayant disparu, les fichiers ont fait l'objet d'un nettoyage assidu.

4.1.2 Le nettoyage

Cette recherche sur les collocations implique un besoin en matériau textuel uniquement ; c'est ce qui nous a guidés dans la phase de nettoyage du corpus TAL(N).

Les informations de mise en page du texte (gras, italique, retrait...) ont été perdues lors de la conversion au format .txt. Nous n'avons pas cherché à les restituer, de même que les informations véhiculées par l'aspect *typo-dispositionnel* [Condamines 2005, p. 18], au demeurant porteuses de sens [Bachimont 2005].

Les en-têtes des articles insérés dans le corps du texte ont également été modifiés après conversion. Nous avons conservé une occurrence du titre et du nom de l'auteur ou des auteurs. Nous avons retiré les titres, les noms d'auteurs et les numéros de page qui figuraient comme rappel en haut à gauche et à droite de la page.

En début d'article, nous avons choisi de supprimer les adresses des laboratoires auxquels appartiennent les auteurs et leurs adresses électroniques.

Dans le corps de l'article, nous avons retiré les notes de bas de page et les numéros associés dans le texte, les extraits de textes encadrés et leur légende, les formules mathématiques, les caractères spéciaux, les graphiques (légende et titre), les diagrammes en bâtons (légende et titre), les tables, les schémas et les exemples.

En fin d'article, nous avons enlevé les dates de soumission et d'acceptation, les références bibliographiques, les remerciements et les annexes.

Nous avons inséré une chaîne de caractères /coupure/, chaque fois qu'une information interrompant le fil continu du discours a été insérée :

- pour marquer une interruption liée à la présence d'un tableau ;
- pour marquer une interruption liée à l'insertion d'un exemple ;

Ex : *Un mouton, s'il mange les arbustes, il mange aussi les fleurs ? (TAL, N° 41.2)*

- lorsqu'une entité pouvait être confondue avec quelque chose d'existant dans le langage ;

Ex : *« je vecteur unité canonique » ou en fait je est je et pas du tout le pronom personnel de première personne du singulier. (TAL, N° 41.2)*

- pour marquer l'insertion d'une citation en anglais ;

Ex: *The meaning of a complex expression is a function of the meanings of its parts and of the syntactic rules by which they are combined. (1990 : 318) (TAL, N° 39.1)*

- pour marquer l'insertion d'une citation d'auteurs en français ;

Ex : « *On doit rechercher ce que les mots veulent dire non pas isolément mais pris dans leur contexte.* » (1884 / 1969 : 122) (TAL, N° 39.2)

- pour marquer l'insertion d'un exemple.

Ex : *Devant tant de mauvaise foi, Jean est resté tout interdit.* (TAL, N° 39.1)

Toutefois, les exemples ne sont pas coupés lorsqu'ils sont insérés dans le corps du texte et remarquables.

Ex : *De même qu'un sens donné peut se réaliser au niveau de la phrase par différentes formes syntaxiques et différentes unités lexicales (« Max a acheté ce vélo à Luc » / « Ce vélo a été vendu à Max par Luc »), au niveau d'un texte, il peut être réalisé sous différentes organisations rhétoriques* (TAL, N° 39.2)

Une chaîne de caractères ?? a été insérée chaque fois que nous avons enlevé un caractère n'interrompant pas le fil du discours :

- en remplacement d'un caractère spécial ;
- en remplacement des puces indiquant une liste d'arguments.

4.1.3 Les traitements

Le corpus TAL(N) n'a pas été balisé [Burnard et Sperberg 1996 ; Ide et Véronis 1996] ; en revanche, il a fait l'objet d'un étiquetage morpho-syntaxique et d'une lemmatisation [Véronis 2000]. L'étiquetage a été effectué avec WinBrill. WinBrill, développé par Brill, est un étiqueteur (ou catégoriseur), c'est-à-dire un outil qui affecte automatiquement à chaque mot une étiquette représentant la catégorie grammaticale (partie du discours) à laquelle il appartient [**Ex :** ABR – Abréviation ; ADJ :sg – Adjectif (sauf Participe passé) au singulier ; ADJ :pl – Adjectif (sauf Participe passé) au pluriel ; ADV – Adverbe]. Cet étiqueteur a d'abord été testé sur l'anglais mais, son principe de fonctionnement étant basé sur l'apprentissage, il s'applique à d'autres langues. La version la plus communément utilisée pour le français est celle entraînée par l'INaLF⁹ (Institut national de la langue française¹⁰) [Lecomte et Paroubek 1996]. Pour pouvoir appliquer WinBrill, le texte doit respecter un format d'entrée particulier au niveau des phrases. Le corpus TAL(N), *via* un script développé localement, a donc été découpé de façon à obtenir une phrase par ligne. En outre, chaque signe de ponctuation a été entouré d'une espace. Ce format est obtenu par l'application du script.

La lemmatisation a été effectuée avec Flemm, développé par Namer [Namer 2000]. Flemm est un lemmatiseur (ou analyseur flexionnel du français), c'est-à-dire qu'il assigne à chaque mot une étiquette réduisant ce mot à sa forme canonique, notamment orthographique. Il ne fonctionne que sur des corpus préalablement étiquetés (soit par Brill, soit par Tree-Tagger).

⁹ Étiqueteur de Brill en version libre, disponible à l'adresse
<http://www.inalf.cnrs.fr/WinBrill/winbrill.tele.html>

¹⁰ <http://www.inalf.cnrs.fr/>

Toutefois, à l'usage, toutes ces données ne se sont pas révélées nécessaires à l'analyse sémantico-discursive menée sur les collocations. Loin d'en faciliter les analyses et de les améliorer, les étiquettes brouillent la lisibilité du corpus. Aussi, nous avons préféré poursuivre la consultation du corpus sur des fichiers au format .txt bruts.

4.2 Résultat

Le corpus TAL(N) a été conçu en 2003. Il comporte un million d'occurrences sans la ponctuation. Il réunit 193 articles de recherche scientifique : 69 articles de revue *TAL* et 124 articles de conférence. Il n'y a pas de problème de surreprésentativité dans l'échantillonnage, dans la mesure où les articles de revue sont deux fois plus longs (environ 20/30 pages) que les articles de conférence (au maximum 10 pages). Le corpus TAL(N) est constitué de textes entiers exclusivement rédigés en français et relevant d'un registre unique qualifié de *scientifique*. Les textes appartiennent à une même période consécutive d'une durée égale : les articles des revues *TAL* parues entre 1998 et 2002 inclus (du volume 39(1) au volume 43(3)) et les articles publiés dans les actes de la conférence TALN entre 1999 et 2003 inclus.

Conclusion

Dans ce chapitre, nous avons posé la question des contraintes à retenir pour la constitution d'un corpus spécialisé dédié à une analyse sémantico-discursive. En effet, cette démarche théorique implique un processus de construction du corpus fondé le plus possible sur des critères internes à un domaine de pratique. Dans cette optique, la procédure est double. Elle consiste d'abord à trouver un moyen de délimiter un domaine de pratique, puis à dresser la liste des critères de sélection textuelle garantissant la représentativité des textes par rapport au domaine retenu, et permettant l'émergence d'une cohésion textuelle. Nous avons choisi pour ce faire le modèle des communautés de discours. Les arguments en faveur de cette théorie ont trait à la délimitation de la communauté de *l'intérieur*, sans recourir à des critères plus subjectifs liés au registre. En outre, cette démarche désigne objectivement plusieurs sources potentielles de données textuelles comme étant représentatives de la communauté. En nous appuyant sur les critères traditionnels de constitution de corpus, cumulés à la fiabilité des critères induits par le modèle de la théorie des communautés de discours, nous avons pu minimiser les effets subjectifs des critères externes de sélection des textes restant à notre charge. La mise en application de ces critères nous a permis de circonscrire la communauté TAL et d'accéder légitimement aux ressources textuelles identifiant le domaine, ressources à partir desquelles nous avons pu constituer le corpus TAL(N) représentatif de la langue de spécialité TAL.

Les contraintes de constitution du corpus électronique spécialisé étant principalement fonction de l'hypothèse de recherche qui en motive le besoin, le chapitre 4, intitulé « Hypothèses de recherche et protocole de validation », en présente la substance.

Chapitre 4

Hypothèses de recherche et protocole de validation

Présentation

Dans les trois chapitres précédents, nous avons défini les collocations textuelles, phénomènes étudiés dans cette recherche, et introduit la problématique de la motivation sémantique sous-jacente à l'association des deux unités lexicales formant la collocation. Nous avons également présenté le cadre théorique de la sémantique des possibles argumentatifs et justifié l'intérêt de le choisir pour éprouver cette problématique dans une perspective argumentative. A dessein, nous avons construit le corpus TAL(N), objet d'étude nécessaire à la validation de cette problématique. Ce faisant, il reste à étayer cette problématique et à se doter d'une méthodologie adaptée à sa validation. Dans cette optique, le chapitre 4 est composé de deux parties. Il expose d'abord le paradigme d'hypothèses constitué à partir du rapprochement des caractéristiques inhérentes aux collocations textuelles et aux possibles argumentatifs ; ces hypothèses sont introduites dans un ordre allant du général au particulier. Il détaille ensuite le protocole de validation propre à tester ces hypothèses. Les ajustements pratiqués au niveau méthodologique se sont traduits par la création d'un processus de validation pluridisciplinaire mêlant linguistique de corpus et lexicographie, adapté à la nature électronique des données et aux contraintes lexicales liées à la sémantique des possibles argumentatifs.

1 Hypothèses de recherche

1.1 Première hypothèse – H1

Parallèlement aux besoins en données textuelles nécessaires à la validation de ce qui a été en premier lieu une orientation de recherche, l'idée de constituer un corpus d'orientation scientifique est née d'un constat dont nous avons pleinement conscience qu'il est stéréotypé, culturellement ancré et qui postule que le discours scientifique spécialisé – DSS – se présente comme le discours de la connaissance, le discours de la construction du savoir sur le monde. Il est courant de trouver dans la littérature consacrée au discours scientifique et/ou à la vulgarisation scientifique des assertions du type « la recherche bâtit une représentation du monde supposée correspondre à la réalité, telle du moins que la définit la collectivité scientifique à un moment donné » [Mortureux 1985] ou encore « la science s'avère d'autant plus péremptoire que son objet paraît restreint » ; « [le discours scientifique tend à] enfermer "sa" vérité dans un champ bien délimité » [Beaune 1988]. Les expressions « une représentation du monde », « la réalité », « péremptoire » et « sa vérité » confortent ce point de vue et poussent à considérer le DSS comme étant essentiellement informatif. Dès lors, il serait dégagé des marques traditionnelles de subjectivité : modalités d'énonciation et modalités d'énoncé. Transposé au modèle de la sémantique des possibles argumentatifs – SPA, les unités lexicales mobilisées dans le DSS ne porteraient pas de valeur axiologique positive ou négative, théoriquement associée à la signification et/ou au sens des mots. Toutefois, ce précepte va à l'encontre d'un principe théorique fondateur en SPA, puisque ce modèle considère la langue générale comme étant par essence évaluative, c'est-à-dire porteuse de visée subjective par nature, eu égard à son fonctionnement profondément argumentatif. Or, bien que singulier par ses préférences lexico-syntaxiques, le DSS recourt à la langue générale pour se construire.

La première hypothèse – H1 – stipule donc que le DSS recèle des marques linguistiques subjectives. Cette hypothèse est volontairement triviale, puisqu'elle a impulsé la constitution du corpus TAL(N) et qu'il s'agissait, autant que faire ce peut, d'éviter l'écueil de la circularité des contraintes dans l'élaboration du corpus (*cf.* Chapitre 3). Par conséquent, il est entendu que répondre par l'affirmative ou la négative présente bien peu d'intérêt ; ce qu'il nous importe de savoir, c'est à quel niveau s'inscrit cette subjectivité et à quels mécanismes elle obéit. Ce qui nous amène à formuler une deuxième hypothèse – H2.

1.2 Deuxième hypothèse – H2

Nous qualifions cette deuxième hypothèse de générale, étant donné qu'elle préexiste à l'observation des données du corpus et fonde à proprement parler les postulats théoriques de la SPA. Les possibles argumentatifs – PA – sont des entités discursives correspondant à la concrétisation dans le discours de l'association du mot avec l'un des éléments de ses stéréotypes. De par le lien argumentatif posé, les PA sont par essence évaluatifs, orientés vers l'un ou l'autre des pôles axiologiques, positif ou négatif. L'orientation positive ou négative des PA est fonction de la contamination discursive, due soit à l'environnement sémantique, soit à l'environnement cotextuel et/ou contextuel. En ce sens, les PA sont des entités fondamentalement subjectives. Cette deuxième hypothèse pose donc que la subjectivité du DSS est inscrite dans les possibles argumentatifs – PA – des unités lexicales qui le composent.

1.3 Troisième hypothèse – H3

La troisième hypothèse – H3 – constitue l’hypothèse particulière à cette recherche et plus fondamentalement encore, la thèse défendue dans ce travail. Cette hypothèse prend la forme d’une intuition issue du rapprochement des définitions respectives de « collocations textuelles » et « possibles argumentatifs ». En effet, partant de la définition de collocations textuelles au sens contextualiste et néo-contextualiste, entendues comme des relations de sens portées par des mots (pleins) qui tendent à apparaître ensemble dans un contexte textuel donné, et de la définition de possibles argumentatifs, entendus comme l’association prévisible de deux unités lexicales liées par une relation d’argumentation et calculable sur la base des enchaînements linguistiques que la signification lexicale est susceptible de générer en contexte, nous postulons qu’une collocation textuelle correspond potentiellement à la concrétisation d’un possible argumentatif dans le discours.

Cette intuition se décline sous la forme d’une triple hypothèse :

- soit la collocation textuelle repérée est la concrétisation d’un PA dans le discours : dans ce cas, la *collocation* est qualifiée d’*argumentative* ;
- soit la collocation textuelle repérée est argumentativement motivée : dans ce cas, le collocatif entretient une relation argumentative avec la base ;
- soit la collocation textuelle repérée ne correspond ni à la concrétisation d’un PA dans le discours, ni à une association argumentativement motivée : dans ce cas, la *collocation* est qualifiée d’*arbitraire*.

1.3.1 Collocation argumentative – H31

La collocation argumentative correspond au cas où la collocation repérée est la concrétisation d’un PA dans le discours. Nous prenons un exemple fondé sur le mot *colère*¹.

Ex : *colère folle*

L’association lexicale *colère folle* correspond à une séquence constituée de deux lexèmes, condition nécessaire à l’obtention du statut de collocation mais pas suffisante en l’absence d’appréciation statistique. À dessein, pour en évaluer la fréquence, nous avons lancé quelques requêtes sur le Web *via* le moteur de recherche Google (en date du 17/04/06). Les requêtes ont été formulées avec guillemets, de sorte que toute association *éloignée* est exclue :

Colère folle : 580 occurrences sur le Web français ;

Colère folle : 672 occurrences sur le Web francophone ;

Ex : *Elle n’entendit encore rien, naturellement ; alors, saisie tout à coup soit de la colère folle de l’être trompé, déçu, soit de l’épouvante de l’être effaré devant un mystère redoutable, soit de l’impatience furieuse de l’être passionné qui rencontre un obstacle, elle saisit la*

¹ Nous avons choisi le mot *colère* pour son utilisation récurrente en qualité d’exemple dans les questions relatives aux collocations [Grossmann and Tutin, à paraître ; Kahane et Polguère 2001b] et parce qu’il présente également l’avantage d’être analysé dans le cadre SPA [Galatanu 2004]. De ce fait, nous disposons de sa signification lexicale et de quelques-uns des paramètres relatifs à son axiologisation.

pincette de la cheminée et frappa le cartel avec tant de force qu'elle le mit en pièces en une seconde. (<http://perso.wanadoo.fr/hypopolo/didac/nouvelle/berthe.htm>)

Colères folles : 234 occurrences sur le Web français ;

Colères folles : 187 occurrences sur le Web francophone.

Ex : *Des fois on a aussi des colères folles. D'ailleurs, en ce moment, ça sent la poudre. Mes citoyens veulent couper les deux réseaux de fibre optique qui traversent la commune mais sans la desservir. Ils disent que le progrès ne vaut que s'il est partagé.*

(<http://www.humanite.presse.fr/journal/2005-04-04/2005-04-04-459679>)

Pour comparaison, nous avons procédé à l'identique pour *folle colère* :

Folle colère : 171 occurrences sur le Web français ;

Folle colère : 138 occurrences sur le Web francophone ;

Ex : *Ce ratage fut même une évidence pour Dieu qui finit par présenter quelques excuses à l'homme et par promettre de ne plus recommencer sa folle colère : l'effacement de l'homme de la surface de la terre.* (<http://alainhenrigangneux.blog.lemonde.fr/alainhenrigangneux/>)

Folles colères : 38 occurrences sur le Web français ;

Folles colères : 74 occurrences sur le Web francophone.

Ex : *Une certaine bassesse me laisse sans défense, et j'ai beau savoir que c'est le Diable qui est là-dessous : j'ai eu des nuits entières secoué par de folles colères.*

(<http://jm.saliege.com/cahiersdumoulin24.htm>)

Nous concluons de ces quelques données que *colère folle* peut être qualifiée de collocation textuelle, eu égard à sa récurrence, notamment au singulier, et que sa forme canonique correspond à la postposition de l'adjectif par rapport à la base².

Pour nous prononcer sur le statut de PA de la collocation *colère folle*, nous avons besoin de la signification lexicale de la base, en l'occurrence *colère*.

² Initialement, nous avons procédé à cette petite expérimentation avec la collocation *violente colère*, pour laquelle les résultats sont statistiquement plus probants : 9 400 occurrences sur le Web français et 16 800 sur le Web francophone (requête formulée au singulier le 14/04/06, adjectif antéposé par rapport à la base). Toutefois, le stéréotype *violent* étant également présent dans le noyau – 2 (cf. Tableau 1) de *colère*, il nous a paru prudent d'exploiter un autre exemple, pour ne pas obscurcir nos propos.

COLERE			
TRAITS NECESSAIRES DE CATEGORISATION (TNC)			
TYPE GRAMMATICAL		ZONE MODALE	ORIENTATION AXIOLOGIQUE
Nom – Féminin		Affectif – Moral	Axiologique – Négatif
ENTITES SEMANTIQUES			ENTITES DISCURSIVES
NOYAU		STEREOTYPES	POSSIBLES ARGUMENTATIFS
1	Penser à faire mal à soi-même ou à autrui	Donc penser mal... Donc mériter punition... Donc vouloir vengeance...	C. donc tourment intérieur
2	Éprouver émotion-sentiment négatif violent	Donc état psychique mauvais... Donc état physique mauvais... Donc agressivité, violence...	C. donc tremblement C. donc comportement violent

Tableau 1 : Signification lexicale de colère³

Nous rappelons qu'un PA correspond à la concrétisation dans le discours de l'association du mot étudié avec l'un des éléments de ses stéréotypes. On parle de bloc d'argumentation externe, puisque consommé dans le discours. Un bloc d'argumentation externe représente donc l'association, dans une structure argumentative, du mot étudié avec l'un des prédicats identifiés comme faisant partie de son argumentation interne.

Pour revenir à notre exemple, dans la description de la signification lexicale de *colère*, on trouve associé au noyau – 2 (entre autres) le stéréotype *état psychique mauvais*. Or le dictionnaire *Le Grand Robert* définit l'adjectif *fou* par « atteint de désordres, troubles mentaux » ou par « qui agit, qui se comporte d'une façon peu sensée, anormale », ce qui correspond de fait à une forme possible d'*état psychique mauvais*. L'adjectif *fou* est donc considéré comme faisant partie des stéréotypes *envisageables* de *colère*, à partir du prédicat stable *état psychique mauvais*. Figure ci-dessous le déploiement du bloc d'argumentation interne correspondant à l'association de *colère* et de *fou*.

Colère folle : Éprouver émotion-sentiment négatif violent → Donc désordres, troubles mentaux... Donc comportement peu sensé, anormal...

Dans la mesure où l'adjonction du collocatif *fou* à la base *colère* forme avec ses stéréotypes proposés et envisageables un bloc d'argumentation interne (H31.1), les quatre occurrences *colère folle*, *colères folles*, *folle colère* et *folles colères* correspondent à la concrétisation d'un PA dans le discours (H31.2). Cette collocation est prévisible, « calculable » sur la base des PA existants. La collocation est argumentative.

Pour autant, la collocation ne forme pas un bloc d'argumentation externe au sens où aucun connecteur argumentatif n'articule le collocatif à la base. Aussi, la collocation figure de notre point de vue la réalisation de surface d'un bloc d'argumentation externe sous-jacent.

³ Le tableau de la signification lexicale de colère reprend la description de la signification lexicale du mot, telle qu'elle est proposée par Galatanu [Galatanu 2004, p. 220].

Conformément aux postulats théoriques de la SPA-ALD, de par le lien argumentatif posé, les PA sont par essence évaluatifs. Ils s'organisent autour de deux faisceaux orientés respectivement vers l'un ou l'autre des pôles axiologiques, positif ou négatif. Aussi, le troisième paramètre inhérent au paradigme argumentatif pose qu'une collocation argumentative est nécessairement intrinsèquement axiologisée (H31.3).

Ex : *colère folle* → Donc attitude destructrice

Ex : *colère folle* → *Pourtant attitude destructrice

Ex : *colère folle* → *Donc attitude positive

Ex : *colère folle* → Pourtant attitude positive

Seuls les enchaînements normatif « attitude destructrice » et transgressif « attitude positive » sont sémantiquement acceptables, ce qui signifie que le PA est intrinsèquement axiologisé négativement.

1.3.2 Motivation argumentative – H32

La motivation argumentative correspond au cas où la collocation textuelle repérée est argumentativement motivée, au sens où le collocatif entretient une relation argumentative avec la base, sans pour autant former un bloc d'argumentation interne, encore moins un bloc d'argumentation externe. Nous poursuivons la démonstration avec un exemple fondé sur le mot *colère*.

Ex : *colère épouvantable*

L'association lexicale *colère épouvantable* correspond à une séquence constituée de deux lexèmes. Nous avons également cherché à en évaluer le caractère significatif par quelques requêtes lancées sur le Web *via* le moteur de recherche Google (en date du 28/08/06). Les requêtes ont été formulées avec guillemets, de sorte que toute association *éloignée* est exclue :

Colère épouvantable : 278 occurrences sur le Web français ;

Colère épouvantable : 503 occurrences sur le Web francophone ;

Ex : *Rétrospectivement, cet événement qui l'a marquée jusqu'aujourd'hui la plonge dans une colère épouvantable.* (<http://www.humanite.presse.fr/journal/2005-04-09/2005-04-09-460016>)

Colères épouvantables : 71 occurrences sur le Web français ;

Colères épouvantables : 125 occurrences sur le Web francophone.

Ex : *Maxime Loupiot voulait être fleuriste quand il serait grand, ce qui mettait son père dans des colères épouvantables : il voulait faire de lui un chasseur comme l'étaient ses aïeux. Mais le jeune loup était têtu... Finalement, Maxime Loupiot sera parfumeur...*

(<http://crdp.ac-bordeaux.fr/cddp40/AVotreDisposition/Valises/Contenus/valise35.html>)

Pour comparaison, nous avons procédé à l'identique pour *épouvantable colère* :

Epouvantable colère : 27 occurrences sur le Web français ;

Epouvantable colère : 75 occurrences sur le Web francophone ;

Ex : *Madame Faujas rentrait, tenant au bras un gros panier de provisions. Elle se débarrassa vite, en voyant son fils dans cette épouvantable colère. Elle lui prit les bras.*

(<http://bibliotheq.net/emile-zola/la-conquete-de-plassans/page-180.html>)

Epouvantables colères : 4 occurrences sur le Web français ;

Epouvantables colères : 4 occurrences sur le Web francophone (identiques aux précédentes).

Ex : *Personnage sombre de l'ennéade d'Héliopolis, Seth terrorise le panthéon égyptien : quand il abuse de la boisson ou se fait berner lors d'un pari, il éclate en d'épouvantables colères qui provoquent orages, pluies et tremblements de terre.*
(<http://membres.lycos.fr/historienduweb/newpage80.html>)

Nous concluons de ces quelques données que *colère épouvantable* peut être qualifiée de collocation textuelle, eu égard à sa récurrence, notamment au singulier, et que sa forme canonique correspond à la postposition de l'adjectif par rapport à la base.

La motivation argumentative correspond au cas où le collocatif entretient une relation argumentative avec la base, sans pour autant former un bloc d'argumentation interne, c'est-à-dire sans que le collocatif fasse partie des stéréotypes proposés ou envisageables de la base. *Le Grand Robert* définit l'adjectif *épouvantable* par « qui cause ou est de nature à causer de l'épouvante, qui effraie fortement », « inquiétant, très mauvais » et « excessif ». Aucune des définitions proposées ne peut être considérée comme faisant partie des stéréotypes proposés ou envisageables de *colère* (cf. Tableau 1). Pour autant, l'adjonction de l'adjectif *épouvantable* à la base *colère* forme un déploiement argumentatif possible.

Colère épouvantable : Éprouver émotion-sentiment négatif violent → Donc causer ou être de nature à causer de l'épouvante, effrayer fortement... Donc inquiétant, très mauvais... Donc excessif...

Dans la mesure où l'adjonction du collocatif *épouvantable* à la base *colère* forme un déploiement argumentatif possible⁴ (H32.1), les quatre occurrences discursives *colère épouvantable*, *colères épouvantables*, *épouvantable colère* et *épouvantables colères* correspondent à une motivation argumentative (H32.2) permettant de décrire et d'expliquer le principe sous-jacent à l'association des deux entités formant la collocation. Le mécanisme de motivation argumentative revêt un caractère moins prédictible que celui mobilisé par les collocations argumentatives. Toutefois l'émergence de principes associatifs récurrents en fonction de la monovalence (positive ou négative) ou de la bivalence de la base, semble possible.

Conformément aux postulats théoriques de la SPA-ALD, de par le lien argumentatif déployé, ces collocations comportent nécessairement une évaluation, orientée vers l'un ou l'autre des pôles axiologiques, positif ou négatif. Aussi, le troisième paramètre inhérent à la motivation argumentative pose également que la collocation est intrinsèquement axiologisée (H32.3).

Ex : *colère épouvantable* → Donc attitude destructrice

Ex : *colère épouvantable* → *Pour autant attitude destructrice

Ex : *colère épouvantable* → *Donc attitude positive

Ex : *colère épouvantable* → Pour autant attitude positive

Seuls les enchaînements normatif « attitude destructrice » et transgressif « attitude positive » sont sémantiquement acceptables, ce qui signifie que la collocation est intrinsèquement axiologisée négativement.

⁴ De même que pour les collocations argumentatives, en l'absence de connecteur argumentatif articulant le collocatif à la base, nous considérons que la collocation textuelle correspond à la réalisation de surface d'un déploiement argumentatif séquentiel sous-jacent.

1.3.3 Collocation arbitraire – H33

La collocation arbitraire correspond au cas où la collocation repérée ne correspond ni à la concrétisation d'un PA dans le discours, ni à un cas de motivation argumentative. Nous reprenons un exemple fondé sur le mot *colère*.

Ex : *colère ensablée*

L'association lexicale *colère ensablée* correspond là encore à une séquence constituée de deux lexèmes, dont nous avons testé le caractère significatif en lançant des requêtes sur le Web *via* le moteur de recherche Google (en date du 17/04/06). Les requêtes ont également été formulées avec guillemets :

Colère ensablée : 6 occurrences sur le Web français/francophone.

Ex : *C'est des lèvres que surgit le relief, dis que ton foyer ignore le toit et l'ombre, que ta porte est murée dans le vide. Mais tu es Qarmate, siècles de colère ensablée. Les clous de ton sarcophage ont été sertis de prophéties.*

(<http://www.humanite.fr/journal/1991-04-30/1991-04-30-642195>)

Aucune occurrence de la collocation plurielle n'a été trouvée, que la requête soit lancée sur le Web français, francophone ou global. « **ensablée colère* » n'a pas fait l'objet d'une requête, dans la mesure où elle est asémantique. Nous concluons de ces maigres données⁵ que *colère ensablée* peut être qualifiée de collocation et que sa forme canonique correspond à la postposition de l'adjectif par rapport à la base.

Pour statuer sur le caractère arbitraire de la collocation, nous recourons de nouveau à la signification lexicale de *colère* (cf. Tableau 1). *Le Grand Robert* définit le participe passé/adjectif *ensablé* par « recouvrir, remplir, combler de sable ». Or, dans la description de la SL de *colère*, parmi les stéréotypes mentionnés ou envisageables, aucune donnée ne fait état d'un stéréotype correspondant à *ensablé* ou permettant de le prédire (ni au sens propre ni au sens figuré). Le participe passé/adjectif *ensablé* ne peut donc être considéré comme faisant partie des stéréotypes proposés et envisageables de *colère*. Figure ci-dessous le déploiement argumentatif correspondant à l'association de *colère* et de *ensablé*, dans ses formes normative et transgressive.

Colère ensablée : Éprouver émotion-sentiment négatif violent → *Donc recouvert, rempli, comblé de sable...

Colère ensablée : Éprouver émotion-sentiment négatif violent → *Pourtant recouvert, rempli, comblé de sable...

Dans la mesure où l'adjonction du collocatif *ensablé* à la base *colère* ne forme pas avec ses stéréotypes proposés et envisageables un bloc d'argumentation interne ou un déploiement argumentatif possible (H33.1), la collocation *colère ensablée* est qualifiée d'arbitraire (H33.2), au sens où elle ne peut être sémantiquement prédictible ; elle n'a de réalité que discursive.

Par l'absence de lien argumentatif, la collocation arbitraire ne devrait pas être intrinsèquement axiologisée, ce que nous discutons. En effet, sans remettre en cause la

⁵ Nous sommes conscients que la faible quantité d'occurrences trouvées pour *colère ensablée* ne satisfait pas pleinement au critère de récurrence inhérent à la définition même de collocation textuelle, ce qui fait de *colère ensablée* une association lexicale dont le statut collocationnel est possiblement discutable. Toutefois, il nous semble souhaitable, pour la clarté de la démonstration, de poursuivre la présentation des hypothèses avec un paradigme d'exemples cohérent et donc une base collocationnelle commune.

dimension évaluative de la langue générale, nous pensons que, bien que neutralisée sémantiquement, la base conserve son potentiel axiologique (H33.3).

Ex : *colère ensablée* → Donc attitude destructrice

Ex : *colère ensablée* → *Pourtant attitude destructrice

Ex : *colère ensablée* → *Donc attitude positive

Ex : *colère ensablée* → Pourtant attitude positive

De même que pour la collocation argumentative, seuls les enchaînements normatif « attitude destructrice » et transgressif « attitude positive » semblent sémantiquement acceptables, ce qui signifie que la collocation est intrinsèquement axiologisée négativement, sans être intrinsèquement argumentative, mais qu'elle conserve son potentiel axiologique.

1.3.4 Quatrième hypothèse – H4

Enfin, la dernière hypothèse – H4 – découle simultanément du modèle SPA-ALD et des hypothèses formulées en H3. Elle allie les dimensions sémantico-discursive et pragmatico-discursives mobilisées par les six mécanismes d'axiologisation du sens au niveau discursif général cotextuel et/ou contextuel (*cf.* Chapitre 2), dont nous pensons qu'ils sont transposables à la description des relations qui régissent l'association entre la base et le collocatif (H41), relations que nous qualifions de macro-sémantiques⁶. Pour rappel, nous mentionnons à nouveau ces six mécanismes en les accompagnant d'exemples.

Ex : Renforcement : *Elle est belle, donc mariée depuis très longtemps.*

Ex : Activation : *Elle est belle, donc elle plaît.*

Ex : Neutralisation : *Tous les ans il y a une grève générale.*

Ex : Affaiblissement : *Elle est belle, mais intelligente.*

Ex : Interversion : *C'est bon la honte.* (Publicité pour les biscuits Senoble)

Ex : Transgression : *Elle est belle, pourtant célibataire.*

La collocation argumentative correspondant à un bloc d'argumentation interne, son orientation positive ou négative résulte d'un mécanisme sémantico-discursif d'axiologisation du sens. Selon la nature normative (argumentation en *donc*) ou transgressive (argumentation en *pourtant*) du bloc d'argumentation interne déployé, l'orientation axiologique du collocatif est conforme à l'orientation axiologique de la base, générant un mécanisme de *renforcement*, ou inverse à celle-ci, générant un mécanisme d'*affaiblissement* par exemple.

Ex : Colère folle : *fou* étant un monovalent négatif, son orientation axiologique est conforme à celle de la base *colère* ; il s'agit d'un mécanisme de *renforcement* des valeurs axiologiques négatives portées par *colère*.

La motivation argumentative correspondant à un déploiement argumentatif possible mais non « prévu » par la signification lexicale de la base, l'orientation positive ou négative des collocations concernées résulte d'un mécanisme pragmatico-discursif d'axiologisation du sens. Là encore, selon la nature normative ou transgressive du déploiement argumentatif, l'orientation axiologique du collocatif est conforme à l'orientation axiologique de la base,

⁶ Traditionnellement, les mécanismes d'axiologisation sont qualifiés de sémantiques, lorsqu'ils désignent des processus internes à la signification lexicale, intervenant au niveau du mot. Aussi, nous qualifions de *macro-sémantiques* ces mêmes processus internes à la signification lexicale, mais à l'échelle des collocations ; c'est-à-dire qu'ils mobilisent deux unités lexicales.

généralisant un mécanisme d'*activation* ou de *renforcement*, ou inverse à celle-ci, généralisant un mécanisme d'*affaiblissement* par exemple.

Ex : Colère épouvantable : *épouvantable* étant un monovalent négatif, son orientation axiologique est conforme à celle de la base *colère* ; il s'agit d'un mécanisme de *renforcement* des valeurs axiologiques négatives portées par *colère*.

La collocation arbitraire ne correspondant ni à un bloc d'argumentation interne, ni à un déploiement argumentatif « calculable » à partir de la signification lexicale de la base, son orientation positive ou négative résulte également d'un mécanisme pragmatique-discursif d'axiologisation du sens. Elle n'a de réalité que discursive, par la capacité du discours à créer des systèmes de valeurs propres à influencer sur le cinétisme des significations.

Ex : Colère ensablée : *ensablée* étant neutre, l'orientation axiologique négative de la base est *neutralisée* ; ce qui ne remet pas en cause le caractère négatif du potentiel axiologique porté par la base *colère*.

Enfin, d'un point de vue discursif général, nous pensons intuitivement que le DSS ne « joue pas » à l'intervention des valeurs axiologiques, eu égard à un contrat de transparence et de non-ambiguïté dans la transmission du savoir. Nous nous attendons donc à trouver une majorité de potentiels activés, voire renforcés, en particulier pour les collocations arbitraires, mais pas ou peu de potentiels neutralisés ou affaiblis et encore moins de potentiels intervertis ou transgressés (H42).

Ex : Colère ensablée : *C'est des lèvres que surgit le relief, dis que ton foyer ignore le toit et l'ombre, que ta porte est murée dans le vide. Mais tu es Qarmate, siècles de colère ensablée. Les clous de ton sarcophage ont été sertis de prophéties.*

(<http://www.humanite.fr/journal/1991-04-30/1991-04-30-642195>)

La présence dans le cotexte de *ton foyer ignore le toit et l'ombre, ta porte est murée dans le vide, les clous de ton sarcophage* nous permet d'observer que la collocation *colère ensablée* est activée négativement, conformément à l'orientation axiologique première de *colère*.

Pour une bonne évaluation de nos hypothèses, nous statuons pour chaque ensemble d'unités lexicales analysées sur les paradigmes d'hypothèses 3 et 4. Nous ne nous prononçons sur la validité des hypothèses 1 et 2 qu'en synthèse générale du chapitre 5, pour garantir aux propos tenus une certaine pertinence.

2 Protocole de validation

Les ajustements pratiqués au niveau méthodologique pour tester nos hypothèses se sont traduits par la création d'un processus de validation pluridisciplinaire mêlant linguistique de corpus et lexicographie, adapté au caractère électronique des données et aux contraintes lexicales liées au modèle SPA.

2.1 Linguistique de corpus

Du point de vue linguistique de corpus, le protocole de validation de nos hypothèses de recherche se décompose en deux étapes : le repérage des unités lexicales et le dépouillement des données.

2.1.1 Le repérage des unités lexicales

Cette étape vise à établir la liste des unités lexicales formant la base des collocations ultérieurement collectées et analysées. Pour ce faire, il est nécessaire d'explicitier et de justifier la démarche préalable au choix des données.

La première approche choisie pour conduire nos analyses sémantico-discursives (*cf.* Chapitre 5) peut être qualifiée de démarche déductive, au sens que la logique donne à ce terme⁷ ; c'est le précepte théorique préexistant à la constitution du corpus TAL(N) en lien avec la subjectivité du DSS qui a guidé le choix des mots à rechercher. Or la déception a été à la mesure de l'enthousiasme initial ; autant les premières analyses se sont avérées prometteuses, autant les données suivantes se sont révélées infructueuses. Aussi, à la lumière des paramètres qui faisaient défaut à cette première approche, nous avons adopté en définitive une démarche inverse, dite inductive⁸. Bien que cette première démarche soit stérile, il nous semble pourtant nécessaire de revenir sur les postulats qui en sont à l'origine et d'en présenter la teneur, ne serait-ce que parce que, par la résistance qu'elle nous a opposée, évitant de reproduire les mêmes erreurs, elle est à l'origine du succès de la démarche inductive.

2.1.1.1 Démarche déductive

2.1.1.1.1 *Les unités lexicales épistémiques*

Nous avons écrit du DSS, d'une part, qu'il se présente comme le discours de la connaissance, le discours de la construction du savoir sur le monde, d'autre part, dans le cadre de notre première hypothèse, qu'il recèle probablement des marques linguistiques subjectives. Partant de ces deux postulats, il nous a semblé pertinent d'analyser les unités

⁷ Définition du TLFi : type de raisonnement qui conduit de une ou plusieurs propositions dites prémisses à une conclusion « nécessaire », c'est-à-dire inévitable si l'on accepte la règle du jeu.

⁸ Définition du TLFi : type de raisonnement consistant à remonter, par une suite d'opérations cognitives, de données particulières (faits, expériences, énoncés) à des propositions plus générales, de cas particuliers à la loi qui les régit, des effets à la cause, des conséquences au principe, de l'expérience à la théorie.

lexicales liées aux modalités épistémiques [Gardies 1983]⁹. En effet, les modalités épistémiques couvrant les zones modales de la « croyance » et de la « connaissance », nous cherchions à observer les marques d’objectivité *vs* de subjectivité possiblement inscrites dans les mots véhiculant les savoirs discutés dans les articles scientifiques. L’unité lexicale *savoir* étant emblématique de cette zone modale, nous avons débuté nos analyses par lui, en tenant compte de son ambivalence grammaticale (verbe et substantif).

2.1.1.1.2 La collecte par synonymes et ses limites

Partant de l’unité lexicale *savoir*, nous avons établi objectivement une liste de mots appartenant à la zone modale de l’épistémique, en recherchant ses synonymes. Nous avons utilisé pour ce faire *Le Dictionnaire des synonymes*, ressource électronique en ligne accessible à partir du site Web du laboratoire de linguistique de l’université de Caen – le CRISCO¹⁰. Ce *Dictionnaire des synonymes* contient approximativement 49 000 entrées et 396 000 relations synonymiques. La base de départ est constituée de sept dictionnaires classiques (Bailly, Benac, Du Chazaud, Guizot, Lafaye, Larousse et Robert) dont ont été extraites les relations synonymiques¹¹.

En entrant le mot *savoir* dans ce dictionnaire, nous avons obtenu la liste des 46 synonymes (plus exactement parasynonymes) mentionnés ci-après : acquis, apprendre, aptitude, bagage, ballot, capacité, cognition, compétence, comprendre, conception, connaissance, connaissances, connaître, croyance, culture, doctrine, érudition, être au courant, être au fait, être averti, être informé, expérience, formation, génie, gnose, humanisme, imagination, imaginer, instruction, intelligence, lecture, lettre, lettres, littérature, lumière, mémoire, notions, omniscience, posséder, pouvoir, pratique, prendre garde, s’attendre, sagesse, science, voir.

Nous avons ensuite recherché dans le corpus TAL(N) des occurrences des noms au singulier et au pluriel (que nous avons restitués au besoin), puis des verbes (également restitués si nécessaire) pour chacun de ces parasynonymes¹².

Ex : *acquis* (singulier et pluriel), *acquérir* ; *apprentissage*, *apprentissages*, *apprendre*

Cette première confrontation aux données du corpus a permis d’éliminer trois types de données :

- les données non présentes en corpus ;

Ex : *bagage*, *ballot*, *cognition*, *érudition*, *être au courant/au fait*, *gnose*, *humanisme*.

- sur la base d’informations co- ou contextuelles, les cas de synonymies trop éloignées, les données sortant généralement du cadre épistémique requis ;

Ex : *culture*, *doctrine*, *génie*, *imagination*, *lecture*, *littérature*, *mémoire*, *posséder*.

⁹ Pour une étude diachronique sur les modalités, voir [Meunier 1979].

¹⁰ Adresse du CRISCO – <http://elsap1.unicaen.fr/index.html>

Lien vers *Le Dictionnaire des synonymes* – <http://elsap1.unicaen.fr/dicosyn.html>

¹¹ Pour une présentation exhaustive des principes de constitution et de consultation du *Dictionnaire des synonymes*, voir [Eufe et al. 2004]. Le *Cahier du CRISCO* N° 17 est disponible sur le Web au format .pdf à l’adresse <http://elsap1.unicaen.fr/cahiers/cahier17.pdf>.

¹² Dans de rares cas, des problèmes d’homophonie ont émergé, par exemple *savons* (nom) et *savons* (indicatif présent, troisième personne du pluriel). Nous les avons éliminés au moment de l’analyse.

- sur la base d'informations co- ou contextuelles, les cas de polysémie, le mot ayant effectivement été trouvé en corpus, mais pour désigner un autre référent.

Ex : *comprendre* (employé généralement au sens de comporter, compter, englober), *instruction* (employé en qualité de terme du domaine informatique), *lettres* (employé essentiellement au sens de caractères typographiques).

À l'issue de cette première confrontation aux données du corpus, la liste des mots candidats au statut de base collocationnelle s'est trouvée réduite au cinquième des possibilités offertes initialement, parmi lesquelles : savoir (substantif singulier et pluriel), savoir (verbe), acquis (singulier et pluriel), acquérir, apprentissage, apprentissages, apprendre, aptitude, être apte, capacité, capacités, être capable, compétence, compétences, compréhension, comprendre, connaissance, connaissances, connaître, expérience, expériences, expérimenter.

Bien que restreinte, cette liste de candidats suffisait potentiellement à nos besoins, dans la mesure où ce qui nous importait n'était pas tant la diversité des bases collocationnelles que la variété des candidats collocations. Mais deux types de difficultés sont apparus :

- d'une part, à l'exception de *connaître* et de *connaissance(s)*, chaque mot pris individuellement a généré peu de candidats collocations ;
- d'autre part, peu de candidats collocations ont ensuite résisté aux quatre critères constitutifs du dépouillement des données.

Ex : *savoir* (nom singulier-pluriel et verbe à l'infinitif) – 256 occurrences en corpus.

→ 4 collocations : *difficile de savoir*, *intéressant de savoir*, *permettre de savoir*, *falloir savoir*.

→ 19 occurrences : uniquement sur le verbe, le collocatif en position tête.

Ex : *capacité(s)* – 95 occurrences singulier et 49 occurrences plurielles.

(*être*) *capable(s)* – 113 occurrences.

→ 3 collocations : *capacité du système*, *capacités de généralisation*, *capacités inférentielles*.

→ 18 occurrences : uniquement sur le nom, la base en position tête.

2.1.1.1.3 Évaluation de la démarche déductive

Cette démarche déductive a généré deux problèmes majeurs pour le repérage et la collecte des unités lexicales propres à servir de base aux candidats collocations visés pour l'analyse sémantico-discursive : un manque affligeant de matière sur laquelle travailler et un manque de « pertinence » des données pour tester les hypothèses posées.

Pour remédier au manque de données, l'expérience aidant, il est apparu évident que partir des données existantes en corpus permettrait d'éviter un certain nombre d'écueils. Premièrement, celui d'avoir à chercher des données inexistantes ; sans qu'elles soient impossibles dans l'absolu, elles ne correspondent à aucune réalité dans TAL(N).

Ex : *bagage* ou *ballot* en tant que synonymes de *savoir*.

Deuxièmement et de façon plus fondamentale, partir du corpus permet de contrôler la profusion initiale des données, dont nous avons observé qu'elle était nécessaire à l'obtention d'un nombre suffisant de collocations après la tâche de dépouillement des données :

Ex : *connaissance(s)* – 190 occurrences au singulier et 533 occurrences plurielles.

→ 17 collocations :

Ex : *connaissance approfondie, connaissances encyclopédiques, bonne connaissance.*

→ 181 occurrences : principalement sur le nom, la base ou le collocatif en position tête.

Enfin, observer préalablement le corpus permet de construire un paradigme cohérent d'unités lexicales représentatives des thématiques et préoccupations scientifiques de la communauté TAL(N) à laquelle nous avons choisi de nous intéresser.

Concernant la qualité des données pour tester les hypothèses posées, le cheminement est plus subtil. En effet, en dehors de toute considération quantitative, les analyses menées sur les bases *connaissance(s)* et *connaître* ont été concluantes [Dubreil 2005a, 2005b]. Sans nous permettre de statuer définitivement sur la vraisemblance des hypothèses posées, par l'absence de collocation arbitraire, elles en ont confirmé l'intérêt. En effet, les résultats ont montré l'existence possible de collocations argumentatives épistémiques. Dans ce cas, la collocation est bien intrinsèquement axiologisée, ici positivement conformément à la polarité de connaissances. D'un point de vue technique, nous avons conclu de cette série d'analyses à la validité des quatre critères participant du dépouillement des données (*cf.* Section 2.1.2.2). D'un point de vue empirique, ces analyses ont permis de clarifier un paramètre important sur la nature des données à collecter. De fait, l'adéquation entre les données recueillies et les hypothèses posées ne repose pas tant sur l'appartenance des bases collocationnelles à la zone modale de l'épistémique que sur leurs valeurs axiologiques, en l'occurrence monovalente positive pour *connaissance(s)* et *connaître*. Cette considération mise à jour, de nouvelles perspectives d'analyses fondées sur une démarche inductive ont émergé.

2.1.1.2 Démarche inductive

2.1.1.2.1 La collecte par liste de fréquence

Nous avons justifié, dans la section précédente consacrée à la démarche déductive, l'intérêt sinon la nécessité qu'il y a à partir des données du corpus pour constituer le paradigme des bases collocationnelles.

Pour consulter les données du corpus, nous utilisons le logiciel WordSmith Tools (version 3.0)¹³, développé par Mike Scott depuis 1996. WordSmith Tools est un logiciel d'analyse lexicale ; de ce fait, il intègre un grand nombre de fonctionnalités (fréquences, concordances, mots clés...) permettant de faciliter l'analyse du comportement des mots et de leur utilisation au sein des textes.

Dans un premier temps, nous avons établi la liste de tous les mots du corpus classés par fréquences décroissantes (relatives et absolues). Conformément aux prédictions de la loi de Zipf¹⁴ relative à la distribution de la fréquence des mots dans un texte, les premières occurrences les plus fréquentes correspondent aux mots vides *de* (Fréq. 55 077), *la* (Fréq. 30 278), *les* (Fréq. 22 235), suivies de loin par les premiers mots pleins tels *corpus* (Fréq. 2 468), *exemple* (Fréq. 1 964) et *système* (Fréq. 1 870), par exemple.

¹³ Logiciel téléchargeable en version 4.0 limitée, à l'adresse <http://www.lexically.net/wordsmith/index.html>

¹⁴ Pour une présentation vulgarisée de la loi de Zipf, voir : http://fr.wikipedia.org/wiki/Loi_de_Zipf. Pour une expérimentation sur corpus réel, voir : <http://users.info.unicaen.fr/~giguette/java/zipf.html>

Ensuite, partant de cette liste générale, pour obtenir une liste de candidats intrinsèquement axiologisés nous avons opéré un nettoyage drastique, en retirant :

- tous les mots vides ;

Ex : *des, le, et, à, en, dans.*

- toutes les occurrences des verbes auxiliaires *être* et *avoir* ;
- toutes les lexies visibles ;

Ex : *du fait, d'une part, eu égard à.*

- tous les noms propres ;

Ex : *Habert, Wordnet.*

- tous mots pleins dont la fréquence absolue était inférieure à 30 occurrences.

Ex : *estimation* (27), *preuve* (16), *hésitation* (8), *renoncement* (4), *introspection* (1).

La liste de mots restants à l'issue de ce nettoyage a fait l'objet d'une attention particulière, puisque seuls les mots intrinsèquement axiologisés positivement ou négativement et les bivalents ont été conservés.

2.1.1.2.2 Les unités lexicales axiologiques

Une fois la liste des unités lexicales axiologiques les plus fréquentes dressée, pour éviter l'écueil de l'amenuisement des ressources après application des critères de dépouillement des données, nous avons vérifié approximativement en corpus pour chaque mot sa propension à former des candidats collocations.

Pour juger des « potentiels collocationnels » des mots axiologisés, nous utilisons le logiciel Contextes (version 1.1.0)¹⁵, développé par Jean Véronis depuis 1998. Contextes est un concordancier pour les corpus écrits et oraux, c'est-à-dire un programme qui permet d'afficher chaque occurrence d'un mot ou d'une forme donnée avec son contexte d'apparition dans un texte ou un corpus de textes. Il possède de nombreuses options autorisant des recherches complexes et une exportation aisée des résultats. Il permet également l'écoute des fragments sonores correspondant aux contextes dans le cas de corpus oraux transcrits et pourvus d'un alignement son-transcription.

Nous avons généré pour chaque mot sa liste de concordances classées par contexte gauche, puis par contexte droit, de façon à faire émerger les unités lexicales cooccurrent fréquemment (*cf.* Annexe 3 sur la liste des concordances droites et gauches de performance-s). Cette tâche accomplie, ayant à cœur de constituer un paradigme de bases collocationnelles équilibré, nous avons retenu six mots :

- deux monovalents positifs, *amélioration(s)* et *performance(s)* ;

¹⁵ Logiciel téléchargeable en version 1.1.0 limitée à 50 occurrences à l'adresse <http://www.up.univ-mrs.fr/veronis/logiciels/Contextes/index-fr.html>

- deux monovalents négatifs, *erreur(s)* et *problème(s)* ;
- deux bivalents, *évaluation(s)* et *résultat(s)*.

Lorsque c'est possible, nous avons volontairement recherché le verbe associé, en l'occurrence *améliorer* et *évaluer*. En effet, nous considérons par exemple que *amélioration sensible*, *améliorer de manière sensible* et *améliorer sensiblement* représentent trois variantes d'une même collocation. Cette démarche permet donc de regrouper plus facilement les collocations entre elles.

Enfin, une fois les candidats collocations recueillis, une phase de dépouillement des données est nécessaire, en vue de collecter non plus des candidats collocations, mais de réelles collocations textuelles.

2.1.2 Dépouillement des données

Pour le dépouillement des données, il convient d'établir une série de critères fiables et stables pour juger des données correspondant à de vraies collocations. Cette sélection s'effectue donc selon quatre critères : un critère statistique, un critère de pertinence, un critère lexical et un critère terminologique.

2.1.2.1 Le critère statistique

Nous avons restreint arbitrairement le critère statistique à la cooccurrence de deux mots apparaissant au moins trois fois dans le corpus, issus de trois articles différents et employés par au moins trois auteurs. Ces trois paramètres devant être simultanément remplis, ils garantissent l'authenticité des collocations retenues, en étant représentatifs des usages linguistiques de la communauté TAL.

Tous les candidats peu fréquents ont donc été retirés.

Ex : *savoir encyclopédique* (Fréq. 2), *savoir clairement* (Fréq. 1), *connaissance détaillée* (Fréq. 2), *connaissances statistiques* (Fréq. 2).

2.1.2.2 Le critère de pertinence

Le critère de pertinence exclut les collocations grammaticales, constituées d'un mot plein suivi d'une unité subordonnée.

Ex : Nom + Prép. → *sentiment envers*.

Ex : Verbe + Prép. → *s'abstenir de*.

Ex : Adj. + Prép. → *absent de*.

De fait, les hypothèses du paradigme – H3 – portent sur les collocations lexicales, les seules à pouvoir faire l'objet d'une description de la signification lexicale propre à former des blocs d'argumentations entre la base et le collocatif.

Tous les candidats composés de mots grammaticaux ont donc été retirés.

Ex : *savoir si/le/ce*, *connaissance(s) de/du/des*.

2.1.2.3 Le critère lexical

Le critère lexical exclut tous les phénomènes phraséologiques figés (expression, locution et proverbe) faisant l'objet d'une entrée lexicographique à part entière et/ou d'une définition particulière. Ces entités ne partageant pas la même signification lexicale que le mot étudié, elles ne peuvent se soumettre aux mêmes déploiements argumentatifs.

Ex : *à savoir* → *Grand Robert* (loc. conj. de coordination) : *c'est-à-dire*.

Ex : *à ma/notre connaissance* → *Grand Robert* (loc. adv.) : autant que je sache, à la connaissance de tous. - Su (au vu et au su).

2.1.2.4 Le critère terminologique

De même que le critère lexical, le critère lexicographique exclut les termes complexes associés à un domaine de pratique, particulièrement l'informatique, la linguistique ou le TAL. Tous les candidats collocations ont fait l'objet d'une vérification systématique. Nous utilisons pour ce faire deux dictionnaires électroniques accessibles sur le Web : *Eurodicautom*¹⁶ et *Le Grand dictionnaire terminologique – GDT*¹⁷.

Le candidat est exclu dès lors qu'il figure dans l'un ou l'autre ou les deux dictionnaires.

Ex : *apprentissage automatique* → *Eurodicautom* et *GDT* : *processus par lequel un ordinateur accroît ses connaissances et modifie son comportement à la suite de ses expériences et de ses actes passés*.

Ex : *extraction de connaissances* → *GDT* : *activité qui consiste à rechercher, non pas la confirmation de l'information connue qui est contenue dans des sources d'information, mais plus encore des perspectives inexplorées et des corrélations nouvelles qui ajoutent au savoir et à la connaissance humaine sur un sujet particulier*.

Parallèlement, en cas de doute nous recourons à la compétence linguistique des locuteurs de la communauté, qui se prêtent au jeu de l'expertise. Par exemple, aucun des deux dictionnaires ne proposent d'acception spécialisée au domaine du TAL pour *savoir-faire*.

Ex : *savoir-faire* → Selon les experts : *savoir-faire* est employé comme traduction littérale du terme anglais *know-how*, synonyme de *recette*, terme spécialisé dans le domaine de la modélisation du dialogue fondée sur la structure intentionnelle.

Nous considérons également que les variantes d'un terme complexe lui sont apparentées. Nous ne les avons pas traitées comme des collocations, bien que satisfaisant aux critères définitoires retenus.

Ex : *acquisition automatique*, *acquisition de connaissances*.

Ces unités lexicales ne présentent aucune ambiguïté, il s'agit de termes mentionnés dans les deux dictionnaires. Mais le corpus TAL(N) comporte également les associations lexicales *acquérir automatiquement* et *acquérir des connaissances*.

Or, lorsqu'on lance une recherche sur le *GDT*, le résultat varie parfois légèrement. Par exemple, pour *apprentissage inductif*, le dictionnaire propose *apprentissage par induction*, traduit en anglais par *learning by induction*.

¹⁶ Dictionnaire consultable à l'adresse <http://europa.eu.int/eurodicautom/Controller>

¹⁷ Dictionnaire consultable à l'adresse

http://www.granddictionnaire.com/btml/fra/r_motclef/index1024_1.asp

Suivant la démarche du *GDT*, tous les candidats correspondant à une entrée dans les dictionnaires utilisés ou dérivant des entrées proposées sont donc laissés de côté pour l'analyse.

Ex : *acquisition automatique de connaissances, acquisition de connaissances, acquérir des connaissances automatiquement, acquérir des connaissances.*

2.2 Lexicographie

Le protocole mis en place pour tester les hypothèses de recherche comporte deux volets : un volet linguistique de corpus et un volet lexicographique.

Le volet linguistique de corpus est indispensable au choix des unités lexicales formant les collocations, tandis que le volet lexicographique est essentiel à la construction de la signification lexicale de ces mêmes unités ; la signification lexicale étant l'outil nécessaire au déploiement des blocs d'argumentations permettant en définitive de statuer sur le caractère argumentatif, motivé ou arbitraire des collocations retenues.

Concernant l'aspect lexicographique, pour la construction de la signification lexicale, le processus consiste à prendre les définitions d'un mot dans deux dictionnaires différents. La difficulté réside dans la « comparabilité » des dictionnaires, c'est-à-dire qu'ils doivent avoir pour objet de description un même état de la langue française.

Conformément aux propos de Rey, « si l'on s'en tient aux dictionnaires incluant une vaste nomenclature, un traitement détaillé de l'usage des mots et de nombreux exemples de leurs emplois, la langue française dispose actuellement de trois descriptions lexicographiques notables : *Le Robert*, *Le Larousse* et *Le Trésor de la langue française* » [Rey 2002]. Concordant avec l'avis éclairé de lexicographes¹⁸, nous retenons pour cette étude *Le Larousse Lexis* et *Le Grand Robert de la langue française*. Il est à noter que les dictionnaires choisis, parus en 1998 et 2001, couvrent la période à laquelle ont été produits et publiés les textes rassemblés dans le corpus TAL(N). *Le Trésor de la langue française* a été laissé de côté (dans la procédure de construction de la signification lexicale seulement), car la description qu'il propose de la langue française s'arrête aux années 1960, l'énorme corpus littéraire rassemblé par le CNRS n'allant pas au-delà (eu égard aux droits d'auteur).

2.2.1 Les dictionnaires Larousse Lexis et Grand Robert

2.2.1.1 Objectif commun

Un dictionnaire est la présentation dans un certain ordre du lexique de la langue.

Le Larousse Lexis – LL – « s'attache à décrire le lexique du français dans sa plus grande extension » [Larousse Lexis 1998].

Le Grand Robert – GR – considère que la langue en général, et le français en particulier, doit être non pas une abstraction figée mais un ensemble d'usages sociaux dans le temps, dans l'espace, dans la réalité humaine et dont la variété reflète celle des groupes sociaux. C'est pourquoi il se donne pour tâche « la richesse d'information dans l'économie de présentation, la modernité sur fond historique, la simplicité de l'exposé pour maîtriser la complexité des faits décrits ; en un mot, la communication d'une image sociale : celle de la culture francophone classique et contemporaine à travers le kaléidoscope des mots » [Rey 2002].

¹⁸ Nous remercions à ce titre Nathalie Gasiglia et Geoffrey Williams.

Le lexique recensé dans le LL et le GR comporte le vocabulaire courant, le vocabulaire des sciences et des techniques, les locutions et syntagmes figés, la partie classique et littéraire, les néologismes, le français marginal, les régionalismes, les emprunts et les affixes.

Le LL et le GR sont tous deux des dictionnaires de langue. À ce titre, ils s'accordent sur « la mission » de la définition, à laquelle ils prêtent deux fonctions principales :

- d'une part, comme dans tout autre dictionnaire, rendre plus claires et plus précises pour le lecteur les formes du lexique et de la phraséologie ;
- d'autre part, rendre compte de la structuration sémantique du lexique et notamment de ses rapports avec la morphologie. C'est le rôle des « définitions renvois », qui soulignent l'unité des familles de mots et paraissent n'apporter qu'une information dérisoire.

Ex : *rangement* : action de ranger.

2.2.1.2 Traditions différentes

2.2.1.2.1 Rappel historique¹⁹

La tradition du dictionnaire de langue dans la culture française s'intègre à l'histoire des ouvrages de référence, domaine immense qui débute au Moyen Âge et se poursuit aujourd'hui. Au Moyen Âge, on dispose de listes de mots et de glossaires, mais pas de dictionnaire à proprement parler. C'est la Renaissance qui inaugure le genre sous la forme de thésaurus bilingues grec et latin décrits en français. En retard sur l'Italie et l'Espagne, la France ne produit son premier dictionnaire français décrit en latin qu'en 1606, le *Thresor*, de Jean Nicot. Mais le besoin d'un ouvrage purement français et plus normatif se fait sentir, aussi, en fondant l'Académie française, Richelieu lui assigne cette tâche. Se sont alors succédé le premier dictionnaire de l'Académie française en trois volumes (1694), le dictionnaire de Richelet (1680), le dictionnaire de Furetière (1690), puis l'Encyclopédie (XVIII^e), visant la description de la langue et des usages basés sur le français parisien distingué.

Il faut attendre le début du XIX^e pour que prolifèrent les dictionnaires en France, ce qui correspond à une mutation sociale : besoins didactiques nouveaux, démocratisation du savoir, libéralisme, évolution du commerce... Les dictionnaires du français adoptent une description du lexique plus abondante, où dominent les points de vue scientifiques et techniques, alors même que le discours est de plus en plus didactique, dans l'intention d'une ouverture sociale. Ces tendances engendrent dans les années 1860 deux grands ouvrages qui auront un effet organisateur essentiel dans le champ du dictionnaire en France. D'un côté, Pierre Larousse tend à développer l'aspect pédagogique démocratique, la terminologie, les connaissances encyclopédiques mises à la portée d'un large public. De l'autre, Émile Littré cherche la description de la langue moderne dans son épaisseur historique, la construction d'une norme fictive voulue scientifique ; c'est l'héritage du GR. Deux tendances, deux traditions différentes.

¹⁹ Quoique moins exhaustif, ce rappel s'inspire largement de l'historique intitulé « La tradition du dictionnaire de langue dans la culture française » proposé par Rey [Rey 2002] dans la préface à la deuxième édition du *Grand Robert de la langue française*.

2.2.1.2.2 *Le Larousse Lexis : un dictionnaire encyclopédique*

Le LL est un dictionnaire encyclopédique, il cherche à rassembler l'ensemble des connaissances et se donne pour objectif de présenter dans chaque article une unité sémantique cohérente, ce qui l'a amené :

- d'une part, à choisir d'énumérer les sens selon l'ordre logique allant du plus courant au plus rare, du général au particulier, du sens de la langue usuelle au sens scientifique et technique ;
- d'autre part, à pratiquer deux opérations complémentaires : le regroupement et le dégroupement.

Le regroupement correspond au rassemblement sous une même entrée des dérivés et des composés du mot, qui par leurs sens se rattachent étroitement à lui. Toutefois, le LL ne cherche pas pour autant à recréer les traditionnelles familles de mots autour d'une racine étymologique commune. À l'intérieur de l'article, les dérivés dont le sens découle du mot souche ne sont pas définis. Cependant, si le dérivé présente une déviation, une restriction ou une extension de sens par rapport au mot souche, la définition est toujours explicitée.

Ex : *grand, grandement, grandet, grandelet, grandissime, grandeur, grandir, agrandir...*

Le dégroupement consiste, pour un mot normalement considéré comme une seule et même unité, à distinguer plusieurs homonymes. De fait, ces mots correspondent généralement à des unités de langue, mais pas à des unités de discours ; ils apparaissent dans des situations d'énonciation différentes et engendrent des séries dérivationnelles différentes.

Ex : *élever* est dégroupé en trois unités de sens, donc trois entrées.

Élever [porter à un degré, niveau supérieur], élévation, élévateur...

Élever [élever des enfants], bien élevé, mal élevé...

Élever [élever des animaux], élevage, éleveur...

2.2.1.2.3 *Le Grand Robert : un dictionnaire alphabétique et analogique*

Le GR est un dictionnaire alphabétique et analogique de la langue française, ce qui l'a amené :

- d'une part, à entrer tous les mots dans l'ordre alphabétique strict. Le GR renonce aux regroupements, qui conduisent souvent à sacrifier les dérivés et composés, au profit du mot vedette. Donc, chaque rubrique correspond à une forme lexicale distincte (nom, adjectif, verbe...) et est illustrée par des exemples littéraires ;
- d'autre part, à rassembler les expressions diverses d'une même « idée » ; il reconstitue sous chaque entrée des champs lexicaux ou sémantiques, de petits univers de discours repérables.

Ex : *facile* renvoie à *aisé, commode, élémentaire, enfantin, simple, faisable...*

En fin d'article figurent les composés et dérivés du mot entré (s'ils y ont donné lieu), mais également les contraires. Ce système de liens sémantiques n'est autre qu'un

hypertexte qui permet pour une sélection de mots apparentés d'avoir accès à des développements homogènes en d'autres lieux de l'ouvrage.

Le système analogique du GR reste toutefois entièrement soumis à sa structure alphabétique.

Le GR peut également être qualifié d'analytique, puisqu'il opère jusqu'à six niveaux d'enchâssement et d'alinéas. Cela correspond par exemple aux pronominaux fréquents et aux participes pouvant fonctionner comme adjectifs, sous un verbe.

Ex : sous le verbe *animer*, on trouve le pronominal *s'animer*, subdiviser en deux sens, et le participe-adjectif *animé-ée*, analysé en six sens.

Les phénomènes de polysémie et d'homonymie obligent à la multiplication des entrées.

Ex : *voler qqn.*, *voler au secours de qqn.*

Si un mot est considéré comme unique, il fait l'objet d'une seule entrée, mais l'article peut alors se décomposer en « paquets » d'emplois, correspondant à des ensembles cohérents.

2.2.1.3 Nomenclature lexicographique

La fixation terminologique d'une nomenclature des dictionnaires est nécessaire au processus de construction de la signification lexicale, puisqu'elle permet de nommer clairement les différentes composantes lexicographiques réutilisées ensuite pour l'élaboration de la signification lexicale des mots.

Le LL et le GR recourent à la même terminologie, quoique plus développée pour le GR. Nous en proposons une synthèse.

2.2.1.3.1 L'article

De physionomie variée, allant de quelques lignes à plusieurs colonnes selon la complexité du mot traité (nom éthique, nom adjectif formé à partir de nom propre, unité lexicale monosémique, unité lexicale polysémique...), il synthétise trois types d'informations différents, reconnaissables à leur typographie distincte :

- les informations métalinguistiques, concernant la langue et les usages (étymologie, données historiques, remarques grammaticales...) ;
- les équivalences synonymiques servant d'explicitations (en particulier les définitions)²⁰ ;
- l'ensemble des extraits de discours, d'énoncés formant l'objet même de la description et illustrant l'emploi des mots (exemples et citations).

²⁰ Chaque dictionnaire mentionne en cours ou en fin d'article les synonymes et les contraires, en fonction du contexte adéquat.

2.2.1.3.2 L'entrée

L'entrée, en capitale grasse dans les deux dictionnaires, est une forme qui correspond à un « mot du dictionnaire ». La première information donnée par une entrée est la graphie du mot, l'arrangement formel de ses lettres. À la graphie est jointe une information systématique sur la prononciation, donnée selon la transcription de l'API (alphabet phonétique international). Viennent ensuite les abréviations correspondant aux caractéristiques grammaticales du mot selon les catégories traditionnelles (nom, verbe, adverbe...).

2.2.1.3.3 La définition

C'est le centre vital de tout dictionnaire de langue, « sa nature est complexe et sa théorie difficile » [Rey 2002]. La définition est une expression destinée à recouvrir exactement et à suggérer ce que les dictionnaires appellent le sens²¹, c'est-à-dire l'ensemble ou l'un des ensembles des valeurs d'emploi d'une suite de sons, de lettres, correspondant à un mot (ou à une locution). La définition est une périphrase synonyme du défini. En ce sens, la définition tend à replacer le mot dans sa construction, à lui associer les mots avec lesquels il peut être employé. Pour les verbes, par exemple, est précisé en premier lieu le type de sujet et/ou de complément qu'ils admettent selon les sens.

Ex : *soigner* : *s'occuper du bien-être et du contentement de (qqn.), du bon état de (qqc.).*

2.2.1.3.4 Les exemples

Les exemples viennent compléter et illustrer la définition. Ils montrent comment le mot s'insère dans le discours. Les contextes, unités de discours où figure l'unité lexicale décrite, et les éléments de contextes cités se divisent en deux catégories : les exemples intégrés au texte et les citations référencées. Cette distinction est valide dans les deux dictionnaires, mais seul le GR l'explicite dans sa préface.

Les exemples intégrés au texte servent à montrer les principales possibilités combinatoires de chaque mot, les assemblages de mots les plus usuels. Ils correspondent soit à des énoncés attestés, soit à des énoncés traités et simplifiés, réduits aux éléments pertinents dans la description. Parmi ces exemples, nombreux sont ceux qui concernent les phénomènes phraséologiques, y compris les collocations.

Les passages sélectionnés pour les citations référencées illustrent un emploi du mot typique ou, à l'inverse, rare. Ils sont généralement extraits de recueils littéraires et didactiques.

²¹ *A contrario*, ce que la SPA nomme la signification.

2.2.2 Principes de construction de la signification lexicale

À présent que nous disposons de deux dictionnaires comparables permettant d'accéder aux définitions des six mots retenus pour l'analyse sémantico-discursive et d'une nomenclature lexicographique propre à nommer les différentes composantes intervenant dans la signification lexicale – SL –, nous en exposons les principes de constitution²².

2.2.2.1 Les traits nécessaires de catégorisation

Les traits nécessaires de catégorisation – TNC – rassemblent trois types d'informations : le type grammatical, la ou les zones modales et l'orientation axiologique du mot.

Le type grammatical est fourni par les informations données dans l'entrée.

savoir : nom – masculin.

La ou les zones modales dont fait partie le mot reposent sur des informations disséminées au sein de l'article : dans la ou les définitions, les exemples et/ou les synonymes (voire antonymes) proposés, dont on compare la teneur avec une liste des valeurs modales [Galatanu 2002c, pp. 17-32].

Ex : *savoir* : Définition 1 : *Ce que l'on sait* ; Synonymes : *culture* ou encore *érudition*.

Ces données permettent de conclure que *savoir* appartient aux valeurs dites du « jugement de vérité » (le certain, l'incertain, l'exclu et le probable) et plus particulièrement à la zone modale de l'épistémique, dédiée précisément au « SAVOIR » (non à la zone du doxologique dédiée au « CROIRE »).

L'orientation axiologique du mot est fonction des mêmes éléments lexicographiques que pour la ou les zones modales.

Ex : *savoir* : Synonymes : *connaissances* ou encore *instruction*.

Ces données permettent de conclure que *savoir* est un monovalent axiologisé positivement.

2.2.2.2 Le noyau

Le noyau étant composé de prédicats abstraits qui rassemblent des primitifs sémantiques invariants [Putnam 1994], nous ne retenons pour la construction du noyau que les définitions communes aux différentes entrées. En outre, toutes les définitions spécialisées à un domaine de la connaissance ou à une communauté professionnelle sont exclues²³. Par ailleurs, nous générons autant de noyaux que de « paires de définitions » différentes.

Pour exemplifier cette démarche, nous présentons les reproductions des articles consacrés à l'entrée *savoir* (en qualité de nom) dans le LL et le GR et nous en construisons le noyau.

²² Pour éclairer les principes de constitution de la SL, nous utilisons le mot *savoir* en tant que nom. Cet exemple est suivi tout au long de cette section, mais la reproduction des articles figurant dans le LL et le GR n'est proposée qu'en section 2.2.2.1.2, intitulée « Le noyau ».

²³ Naturellement, il ne nous a jamais été donné de rencontrer des définitions spécialisées au domaine de l'informatique ou de la linguistique, sans quoi nous aurions eu affaire à un terme.

♦ **savoir** n. m. (842). Ensemble des connaissances acquises par l'étude : *Le savoir qui compte est celui qu'on se donne soi-même, par curiosité naturelle, passion de savoir* (Léautaud). *Il était convaincu qu'il imaginait tout, mettant à contribution à la fois son savoir et son intelligence* (Caillois) [syn. CULTURE, ÉRUDITION, INSTRUCTION].

Tableau 2 : LL : article consacré à l'entrée savoir

Le LL ne propose qu'une définition :

Définition : ensemble des connaissances acquises par l'étude.

Le GR (*cf.* Tableau ci-dessous) propose trois définitions :

- une première définition générale que nous conservons pour la construction du noyau ;
- une deuxième définition restreinte à l'emploi du mot au singulier, que nous excluons pour sa mention restreinte au GR ;
- une troisième définition spécialisée au domaine de la philosophie, que nous excluons précisément pour cette raison.

Définition : ensemble de connaissances assez nombreuses, plus ou moins systématisées, acquises par une activité mentale suivie.

Ces deux définitions montrent que le noyau de *savoir* est résultatif ; nous retenons :

Noyau : Résultat : ensemble des connaissances acquises.

2. SAVOIR [savwaR] n. m. – Fin XII^e, *saveir* ; *savir*

« *sagesse, intelligence* », 842 ; de 1. *savoir*

- ♦ 1 (V. 1155) Ce que l'on sait (I., A., 2.) ; ensemble de connaissances assez nombreuses, plus ou moins systématisées, acquises par une activité mentale suivie. → **Acquis, connaissances(s), culture** (cit. 12), **érudition, lumière(s), science** (→ Onguent, cit. 2, France). *Un grand savoir acquis par l'étude* (→ **Instruction**), *par l'expérience, Grand savoir. Le savoir de qqn. L'étendue de son savoir* (→ Examiner, cit. 16). – *Plein de savoir* (→ Haut, cit. 46). *Pauvre* (cit. 18) *de sens et de savoir, Mettre l'astuce* (cit. 1) *à la place du savoir*. – *Le vrai savoir* (→ Fondement, cit. 5). *Un savoir affecté* (2. Affecter, p. p. adj., cit. 14), *un vain savoir* (→ Pot, cit. 12) – *Le savoir de qqch.* → **Connaissance**.
- 1. Quel grand et vrai savoir des choses de ce monde (...)

A. DE MUSSET, *Poésies nouvelles*,
« Une soirée perdue ».
- 2. Amer savoir, celui qu'on tire du voyage !

BAUDELAIRE, *les Fleurs du mal*,
« Tableaux parisiens », CXXVI, VII.
- 3. (...) je respecte trop la petite part de savoir que je possède, qui m'a coûté tant de peine à acquérir, pour y introduire des éléments douteux.

BERNAROS, *les Grand Cimetières sous la lune*, p. 248.
- 4. À proportion que la science élargit son pouvoir, elle se tient moins assurée de son savoir.

Jean ROSTAND, *Pensées d'un biologiste*, p. 139,
in FOULQUIE, *Dict. de la langue philosophique*,
art. *Science*.

Un, des savoirs. → **Connaissance**.
- 5. Tel que l'on croit complet et maître de toute chose
Ne dit pas les savoirs qu'à tort on lui suppose (...)

A. DE VIGNY, *Poèmes philosophiques*, « La flûte », III.
- ♦ 2 *Le savoir* : les savants, la science.

« *Que la raison conduise et le savoir éclaire* » (cit. 19).
- 6. Le savoir est honteux, depuis que l'ignorance
A versé son venin dans le sein de la France.

Théophile DE VIAU, *Élégie*, « À une dame ».

Loc. *Le gai** (cit. 4) *savoir*.
- ♦ 3 Philos. État de l'esprit qui sait (I., A., 2.) ; connaissance. Relation entre le sujet et l'objet de pensée dont il admet la vérité, pour des raisons intellectuelles et communicables, à la différence de certitude*, croyance*, foi* (cit. 26) ; cet objet de pensée. *Dans la perception* (cit. 8), *un savoir se forme*.

CONTR. **Ignorance**.

Tableau 3 : GR : article consacré à l'entrée savoir

2.2.2.3 Les stéréotypes

Les stéréotypes correspondent à la déclinaison des prédicats abstraits du noyau par association avec d'autres invariants (représentations), dans des blocs de signification argumentatifs posés comme naturels, du type cause-effet, symptôme-phénomènes, but-moyen, résultat-processus... Ils forment des blocs d'argumentations internes (à la signification lexicale) (*cf.* Chapitre 2). Aussi, les stéréotypes sont fournis par certaines informations présentes au niveau de la définition et qui n'ont pas été mobilisées pour la construction du noyau, mais également par les synonymes. Nous poursuivons l'exemple de *savoir* :

Noyau : Résultat : ensemble des connaissances acquises.

Stéréotype : → Donc étude, activité mentale suivie...

Stéréotype : → Donc culture, érudition, instruction...

2.2.2.4 Les possibles argumentatifs

Les possibles argumentatifs correspondent à la concrétisation dans le discours de l'association du mot avec l'un des éléments de ses stéréotypes. Ils forment des blocs d'argumentations externes (à la signification lexicale), c'est-à-dire qu'ils se réalisent dans une structure argumentative, par l'association du mot dont on construit la SL avec l'un des prédicats identifiés comme faisant partie de son argumentation interne (*cf.* Chapitre 2). Aussi, les possibles argumentatifs sont essentiellement fournis par les exemples.

Nous finalisons l'exemple de *savoir* et proposons ci-dessous un tableau de synthèse :

Noyau : Résultat : ensemble des connaissances acquises.

Stéréotype : → Donc étude, activité mentale suivie...

Stéréotype : → Donc culture, érudition, instruction...

Possibles argumentatifs : → Savoir donc longue étude ; Savoir donc apprentissage

Possibles argumentatifs : → Savoir donc culture immense ; Savoir donc grande instruction

SAVOIR(S)		
TRAITS NECESSAIRES DE CATEGORISATION (TNC)		
TYPE GRAMMATICAL	ZONE MODALE	ORIENTATION AXIOLOGIQUE
Nom – Masculin	Épistémique – Axiologique	Monovalent – Positif
ENTITES SEMANTIQUES		ENTITES DISCURSIVES
NOYAU	STEREOTYPES	POSSIBLES ARGUMENTATIFS
Résultat : ensemble des connaissances acquises	Donc étude, activité mentale suivie... Donc culture, érudition, instruction...	S. donc longue étude S. donc apprentissage S. donc culture immense S. donc grande instruction

Tableau 4 : Signification lexicale de savoir

D'un point de vue strictement méthodologique, d'une part, dans la section consacrée à la construction de la SL de *améliorer* – *amélioration(s)* et *évaluer* – *évaluation(s)*, nous proposons en premier lieu la SL du verbe, car les définitions relatives aux noms commencent généralement par « action de + verbe », ce qui permet de progresser plus logiquement ; d'autre part, les définitions du LL étant en nombre plus restreint, les sections consacrées à la construction de la SL commencent par lui, ce qui évite de réécrire inutilement les nombreuses définitions du GR, que nous n'aurions pas à réutiliser.

2.3 Validation des données

Cette section est consacrée aux outils mis en place pour valider le paradigme d'hypothèses H3 présenté en ouverture de ce chapitre.

2.3.1 Le test de substitution argumentatif

Le test de substitution argumentatif – TSA – est dédié au paradigme d'hypothèses H3, plus particulièrement H31.1/2, H32.1/2 et H33.1/2. Il permet de statuer sur le caractère argumentatif, motivé, ou arbitraire des collocations analysées, puisqu'il vise à restituer si possible un bloc d'argumentation interne, ou à défaut, un déploiement argumentatif possible. Concrètement, l'application du TSA consiste à *substituer*²⁴ les représentations sémantiques participant de la signification lexicale du collocatif, au collocatif lui-même. Eu égard aux hypothèses posées, nous avons vu que trois cas peuvent se produire :

- le cas où les prédicats sémantiques de la signification lexicale du collocatif correspondent à au moins l'un des stéréotypes proposés ou envisageables de la base ; alors la collocation est argumentative car le déploiement argumentatif forme un bloc d'argumentation interne (**Ex** : *colère folle* ; cf. Section 1.3.1) ;
- le cas où les prédicats sémantiques de la signification lexicale du collocatif ne correspondent pas à au moins l'un des stéréotypes proposés ou envisageables de la base, mais forment un déploiement argumentatif possible ; alors la collocation est argumentativement motivée (**Ex** : *colère épouvantable* ; cf. Section 1.3.2) ;
- le cas où les prédicats sémantiques de la signification lexicale du collocatif ne correspondent ni à au moins l'un des stéréotypes proposés ou envisageables de la base, ni à un déploiement argumentatif possible ; alors la collocation est arbitraire (**Ex** : *colère ensablée* ; cf. Section 1.3.3) ;

Pour opérer la substitution des prédicats sémantiques de la signification lexicale au collocatif, nous utilisons deux dictionnaires : *Le Grand Robert* électronique pour sa facilité de

²⁴ Le test de substitution argumentatif doit son nom à ce processus.

consultation et occasionnellement le TLFi²⁵. Ce test est appliqué en tenant compte des variations de noyau entre le nom et le verbe²⁶.

Ex : *améliorer la cohérence* → rendre meilleur, devenir plus satisfaisant → Donc harmonieux, logique, moins contradictoire...

Lorsque la collocation textuelle présente plusieurs formes, nous avons tenu compte des variations éventuelles des éléments de la signification lexicale du collocatif.

Ex : *amélioration sensible* → rendre meilleur, plus satisfaisant → Donc grand, intense, important, manifeste...

Ex : *améliorer sensiblement* → rendre meilleur, devenir plus satisfaisant → Donc changer d'une manière appréciable, importante, notable...

Enfin, nous avons choisi de ne pas accorder les éléments de signification du collocatif en fonction du collocatif lui-même. Dans l'exemple ci-dessous, *performances* est au pluriel, mais *exploit* et *réussite exceptionnelle* restent au singulier.

Ex : *amélioration des performances* → rendre meilleur, plus satisfaisant → Donc exploit, réussite exceptionnelle... Donc rendement maximal, fiabilité élevée...

Le test de substitution argumentatif est appliqué systématiquement à toutes les collocations. Les enchaînements argumentatifs sont d'abord déployés dans leur forme normative, puis transgressive lorsque le premier enchaînement est asémantique. Nous attribuons le statut de collocation arbitraire à toute collocation pour laquelle les enchaînements normatif et transgressif génèrent des déploiements argumentatifs asémantiques.

2.3.2 Le test d'enchaînement argumentatif

Le test d'enchaînement argumentatif – TEA – est également dédié au paradigme d'hypothèses H3, plus précisément H31.3, H32.3 et H33.3. Il permet de statuer sur la réalité de l'axiologisation intrinsèque des collocations, notamment arbitraires, et de définir l'orientation axiologique, positive ou négative, des collocations analysées. En effet, nous considérons qu'une collocation est intrinsèquement axiologisée dès lors qu'au moins un enchaînement normatif et un enchaînement transgressif déployés autour d'antonymes sont sémantiques. Les enchaînements argumentatifs proposés sont construits, au sens où nous les avons artificiellement générés. Toutefois, ils s'inspirent des éléments discursifs rencontrés dans le corpus TAL(N).

Ex : *amélioration de la cohérence* → Donc progrès.

Ex : *amélioration de la cohérence* → *Pour autant progrès.

Ex : *amélioration de la cohérence* → *Donc échec.

Ex : *amélioration de la cohérence* → Pour autant échec.

Seuls les enchaînements normatif « progrès » et transgressif « échec » sont sémantiquement acceptables, ce qui signifie que le PA est intrinsèquement axiologisé positivement, en l'absence de tout élément co- ou contextuel.

De même que le TSA présenté ci-dessus, le TEA est appliqué à toutes les collocations analysées.

²⁵ Dictionnaire électronique consultable à l'adresse <http://atilf.atilf.fr/tlf.htm>

²⁶ Les exemples sont tirés des données recueillies et analysées pour la base *amélioration(s)*. Par conséquent, les éléments de la signification lexicale de *amélioration(s)*, ainsi que les analyses respectives de ces collocations textuelles sont reprises au chapitre 5 (cf. Sections 1.1.2 et 1.1.3).

2.3.3 Considérations méthodologiques

2.3.3.1 Les tableaux de données

Toutes les collocations collectées pour une même base collocationnelle sont présentées sous forme de tableaux. Apparaissent en premier lieu les tableaux des collocations basées sur le contexte droit, eu égard au déploiement argumentatif qui s'opère préférentiellement à droite, suivis ensuite des tableaux des collocations basées sur le contexte gauche.

Les tableaux concernant les substantifs rassemblent les collocations construites sur le nom au singulier, auxquelles nous avons ajouté, quand elles sont identiques, leurs homologues pluriels trouvés en corpus.

Ex : *amélioration des résultats* et *améliorations des résultats* sont regroupées.

Aussi, les tableaux concernant les substantifs au pluriel rassemblent uniquement les occurrences de la collocation qui n'apparaissent que dans la forme plurielle.

De même, les collocations relevées dans le contexte droit et celles relevées dans le contexte gauche sont également regroupées lorsqu'elles sont identiques.

Ex : *amélioration de la qualité* et *la qualité doit être améliorée*.

Chaque tableau comporte cinq colonnes.

Dans la première colonne, intitulée « collocations lexicales », les collocations sont classées par ordre alphabétique et sont recensées dans leur forme canonique, ce qui correspond généralement à la fréquence la plus élevée. Toutefois, par souci de transparence, nous avons parfois restitué la forme canonique de la collocation, alors qu'elle n'était pas présente en corpus (*cf.* Exemple ci-dessous).

La deuxième colonne correspond à la fréquence absolue de la collocation dans sa forme canonique. Lorsque l'on trouve le symbole « - », cela signifie que la collocation lexicale est présente non pas dans sa forme canonique, mais dans une forme apparentée.

Ex : *performance des méthodes* « - », pour *performance de ces méthodes*.

La troisième colonne, intitulée « autres formes », rassemble toutes les variantes présentes en corpus de la collocation mentionnée en première colonne. Y figurent en premier lieu les variantes de la collocation au singulier, suivies de la forme plurielle et de ses variantes, puis de la forme verbale et de ses variantes, si elles existent. Nous entendons par variantes :

- les occurrences directes, mais à cooccurrences distantes ;

Ex : pour *amélioration des performances* → *amélioration sensible des performances*.

Elles sont abrégées comme suit : -A. [...] des performances ;

- les occurrences que nous avons jugées équivalentes, de cooccurrences immédiates et distantes ;

Ex : pour *amélioration des systèmes* → *amélioration du système*.

Ex : pour *amélioration des systèmes* → *amélioration générale de leur système*.

Les collocations lexicales immédiates sont notées dans leur intégralité, tandis que les collocations distantes sont également abrégées ainsi : -A. [...] système.

La quatrième colonne indique les fréquences absolues respectives de chacune des variantes de la collocation lexicale traitée.

Enfin, la cinquième colonne totalise les fréquences absolues de toutes les occurrences.

2.3.3.2 Synthèse

Cette synthèse concerne les considérations méthodologiques, puisqu'elle vise à retracer le « parcours de validation » soumis à chacune des six unités lexicales retenues (*amélioration(s)*, *performance(s)*, *erreur(s)*...) et aux collocations qui leur sont associées.

Chaque base fait l'objet d'un dépouillement des données mentionnant ses fréquences absolues dans le corpus TAL(N), suivies du nettoyage, qui reprend les quatre critères (statistique, pertinence, lexical et terminologique) mentionnés en section « linguistique de corpus » (*cf.* Section 2.1). Cette étape permet d'aboutir à la présentation des listes des collocations recensées sous forme de tableaux.

Vient ensuite la présentation de la construction de la SL de chaque mot, composée des définitions respectives du LL, puis du GR, permettant d'établir le tableau de synthèse sur lequel s'appuient les analyses sémantico-discursives. Nous avons décidé de construire la SL de chaque unité exhaustivement, c'est-à-dire en sachant pertinemment que certains noyaux ne nous serviraient probablement pas.

Ex : Erreur : SL – noyau 3 : action regrettable, inconsidérée, maladroite, déraisonnable.

Les analyses succèdent au tableau de la SL. En vue de dégager des régularités utiles aux conclusions de cette recherche, nous avons réparti les collocations en deux catégories, celles pour lesquelles le collocatif est un nom et celles pour lesquelles le collocatif est un adjectif. Pour la même raison, les collocatifs sont classés en quatre catégories : les monovalents positifs, les monovalents négatifs, les bivalents et les neutres.

Ex : *gain de performance* : monovalent positif.

Ex : *dégradation des performances* : monovalent négatif.

Ex : *performance du modèle* : bivalent.

Ex : *performance de l'algorithme* : neutre.

La présentation des analyses suit ce découpage. Chaque ensemble de collocations est soumis au TSA, puis au TEA, de façon à évaluer les paradigmes H3 et H4. Pour statuer sur l'axiologisation discursive des collocations (H42), aucun test n'est nécessaire, une observation des extraits de discours au niveau co- et contextuel suffit.

Conclusion

Partant de la problématique d'une motivation argumentative sous-jacente à l'association des unités lexicales formant une collocation, le chapitre 4 visait, d'une part à présenter les hypothèses posées pour éprouver cette problématique sur le corpus TAL(N), et d'autre part à proposer un protocole de test adapté.

La progression des hypothèses allant du général au particulier, elle a permis de retracer notre démarche de travail. La première hypothèse détermine en effet l'orientation de la recherche sur la subjectivité du discours scientifique spécialisé, eu égard à la dimension évaluative de la langue générale. La deuxième hypothèse inscrit cette subjectivité dans les possibles argumentatifs des mots que le discours scientifique spécialisé mobilise. La troisième hypothèse fonde la thèse défendue dans ce travail ; elle instancie les trois possibilités offertes par le

rapprochement des entités « collocations textuelles » et « possibles argumentatifs » : soit la collocation textuelle est argumentative, elle correspond à la concrétisation d'un possible argumentatif dans le discours ; soit la collocation textuelle est argumentativement motivée, dans ce cas le collocatif entretient une relation argumentative avec la base ; soit enfin la collocation repérée est arbitraire, elle ne correspond ni à la concrétisation d'un PA dans le discours, ni à une association argumentativement motivée. La quatrième et dernière hypothèse allie les dimensions sémantico-discursive et pragmatico-discursive mobilisées par les six mécanismes discursifs d'axiologisation du sens. Nous pensons d'une part, que ces mécanismes sont transposables à la description des relations de collocation que nous qualifions de macro-sémantiques ; nous pensons d'autre part, que d'une façon générale le discours scientifique spécialisé applique naturellement un contrat de transparence et de non-ambiguïté dans la transmission du savoir, ce que nous pensons observer à travers une majorité de potentiels axiologiques activés, voire renforcés.

La mise en place du protocole de validation des hypothèses suit trois étapes : le repérage des unités lexicales, le dépouillement des données, et la construction de la signification lexicale. Le repérage des unités lexicales a d'abord fait l'objet d'une approche déductive ; les mots à analyser ont été choisis en fonction de la subjectivité supposée du discours scientifique spécialisé. Or, la qualité et la quantité de données récoltées se sont avérées si pauvres, que nous avons adopté en définitive une démarche inductive inverse. L'approche inductive prône un processus de sélection des données fondé sur leur présence significative en corpus et leur pertinence comparée à l'hypothèse de la motivation argumentative posée. Cette pertinence repose sur l'axiologisation positive ou négative des unités sélectionnées. Dès lors, nous avons pu construire un paradigme expérimental équilibré, composé de deux monovalents positifs – *amélioration* et *performance* –, de deux monovalents négatifs – *erreur* et *problème* – et de deux bivalents – *évaluation* et *résultat* –. Pour le dépouillement des données, nous avons rassemblé quatre critères : un critère statistique, un critère de pertinence, un critère lexical et un critère terminologique. La fiabilité du dépouillement tient à l'application simultanée de ces quatre critères à un même mot. Concernant l'aspect lexicographique, pour la construction de la signification lexicale, le processus consiste à prendre les définitions d'un mot dans deux dictionnaires différents, en l'occurrence *Le Grand Robert* et *Le Larousse*. La fixation terminologique d'une nomenclature des concepts et des termes associés dans ces deux dictionnaires est nécessaire au processus de construction de la signification lexicale. En effet, elle permet de nommer clairement les différentes composantes lexicographiques réutilisées ensuite pour l'élaboration de la signification lexicale. Enfin, la mise en place de deux tests, l'un de substitution argumentatif, l'autre d'enchaînement argumentatif, sur les données recueillies dans le corpus TAL(N) à partir des six mots choisis, va permettre de valider ou d'invalidier le principe sémantico-discursif ou pragmatico-discursif de la motivation collocationnelle.

Chapitre 5

Analyse sémantico-discursive des collocations textuelles

Présentation

Le chapitre 5 présente l'ensemble des analyses réalisées sur les six unités lexicales sélectionnées pour former les bases des collocations étudiées. Le chapitre est construit autour de trois grandes sections rassemblant les monovalents positifs *amélioration* et *performance*, les monovalents négatifs *erreur* et *problème* et les bivalents *évaluation* et *résultat*. Chacune de ces sections reprend ensuite la progression établie par la méthodologie et le processus de validation présentés au chapitre précédent : dépouillement des données, construction de la signification lexicale et analyse des données en fonction de la nature grammaticale du collocatif. Puis, l'analyse sémantico-discursive de chaque binôme donne lieu à une synthèse, répondant à chacune des hypothèses posées, établie en fonction de la monovalence positive-négative ou de la bivalence des bases regroupées. Enfin, le chapitre s'achève sur une synthèse générale de l'ensemble des analyses menées sur les régularités observées par type de collocation : collocation argumentative, collocation argumentativement motivée ou collocation arbitraire.

1 Les monovalents positifs

1.1 Améliorer – Amélioration(s)

1.1.1 Dépouillement des données

1.1.1.1 Fréquences et nettoyage

Le corpus TAL(N) contient 185 occurrences du verbe *améliorer* à l’infinitif et 100 occurrences correspondant à différentes formes fléchies, dont la fréquence absolue respective est indiquée entre parenthèses¹ : *amélioraient* (1), *améliorant* (7), *améliore* (30), *amélioré* (6), *améliorée* (26), *améliorées* (5), *améliorent* (12), *améliorerait* (2), *améliorés* (10), *améliorons* (1). On compte également 134 occurrences de *amélioration* au singulier et 40 occurrences du mot au pluriel.

Selon le critère statistique, nous éliminons les candidats trop peu fréquents. Nous retirons donc tous les apax et les cooccurents de fréquence deux².

Ex : *amélioration constante* (1), *améliorations absolues* (2), *améliorer la traduction* (1), *amélioration globale* (3, AC_0117, AC_0112), *améliorer l’analyse* (3, AR_0049).

Selon le critère de pertinence, nous laissons de côté toutes les collocations grammaticales et autres constructions dont le collocatif ne présente pas de contenu sémantique.

Ex : *amélioration de/des/du*, *amélioration par rapport à*, *aucune amélioration* (4).

Aucune lexie apparentée au critère lexical n’a été recueillie, de même qu’aucun terme complexe, construit sur *amélioration(s)* ou *améliorer* et entrant dans la catégorie du critère lexicographique, n’a été relevé dans le corpus.

1.1.1.2 Liste des collocations lexicales

Cette sous-section présente les trois tableaux récapitulatifs des collocations lexicales construites sur *amélioration(s)* et *améliorer*. Les deux premiers tableaux correspondent au contexte droit et le dernier, au contexte gauche du substantif.

¹ Ce principe est valable pour toutes les analyses ; chaque fois qu’un chiffre ou un nombre figure entre parenthèses, il indique la fréquence absolue du mot considéré.

² Une entité (nom et verbe) qui apparaît 459 fois au total dans le corpus génère naturellement une grande quantité d’apax et de cooccurents de fréquence deux. Aussi, nous ne proposons pour le critère statistique que des exemples, choisis parmi les plus significatifs. En outre, nous avons élargi cette démarche à tous les critères intervenant dans le nettoyage des données, lorsqu’elles dépassent cinq occurrences, et à toutes les unités lexicales analysées.

COLLOCATIONS LEXICALES	FREQ.	AUTRES FORMES	FREQ.	TOTAL
Amélioration de la cohérence	1	-Améliorer la cohérence	3	4
Amélioration de la méthode	2	-A. de ces méthodes -A. de méthodes -Améliorer une méthode -cette méthode [...] être A.	1 1 1 1	6
Amélioration de la pertinence	1	-Améliorer la pertinence -Améliorer [...] la pertinence	2 1	4
Amélioration de la qualité	4	-A. [...] de la qualité -Améliorer la qualité -Améliorer [...] la qualité -la qualité [...] être A.	1 10 1 2	18
Amélioration des performances	13	-A. [...] des performances -A. de la performance -Améliorations [...] performance -Améliorer la performance -Améliorer les performances -Améliorer [...] les performances -Améliorer leurs performances -Améliorer notre performance -Améliorer sa performance -les performances [...] être A.	3 1 1 7 38 12 1 2 1 2	81
Amélioration des résultats	12	-A. [...] des résultats -A. du résultat -A. en termes de résultats -Améliorations des résultats -Améliorer ces résultats -Améliorer les résultats -Améliorer [...] les résultats -Améliorer leurs résultats -Améliorer nos résultats -ces résultats [...] être A.	2 1 1 1 1 16 9 1 2 2	48
Amélioration des systèmes	3	-A. du système -A. [...] de leur système -A. de système -Améliorer le système -Améliorer les systèmes -Améliorer nos systèmes -Améliorer un système	1 1 1 2 2 1 1	12
Amélioration sensible	8	-A. [...] sensible -Améliorer de manière sensible -Améliorer sensiblement	2 1 3	14
Amélioration significative	4	-A. [...] significative -Améliorations significatives -Améliorer [...] significative -Améliorer significativement	3 2 2 6	17
Amélioration substantielle	1	-Améliorations substantielles -Améliorer substantiellement	2 1	4

Tableau 1 : Collocations lexicales contexte droit de amélioration(s)

COLLOCATIONS LEXICALES	FREQ.	AUTRES FORMES	FREQ.	TOTAL
Améliorer l'efficacité	3	-A. [...] l'efficacité	1	4

Tableau 2 : Collocations lexicales contexte droit de améliorer

COLLOCATIONS LEXICALES	FREQ.	AUTRES FORMES	FREQ.	TOTAL
Légère amélioration	4	-Légères améliorations	1	5

Tableau 3 : Collocations lexicales contexte gauche de amélioration(s)

1.1.2 Significations lexicales de améliorer – amélioration

1.1.2.1 Signification lexicale de améliorer

Le LL propose une entrée groupée de – améliorable, améliorateur, amélioratif, amélioration, améliorer – sans article associé, mais avec une indication de report à l'entrée *meilleur* (cf. Annexe 4). *améliorer* et *s'améliorer* représentent des sous-entrées dans l'article *meilleur*. Une seule signification est proposée pour chacun d'entre eux.

Définition 1 : le rendre meilleur.

Le LL propose des exemples – *Son caractère ne s'améliore guère* – pour le verbe dans son emploi pronominal.

Définition 1bis : devenir meilleur.

Le GR propose une entrée pour *améliorer* sous laquelle figurent en sous-entrée le verbe pronominal – s'améliorer – et le participe passé/adjectif – amélioré (cf. Annexe 4).

Concernant *améliorer*, l'article se décompose en trois significations : une première signification générale, que nous conservons pour la construction de la SL, une deuxième signification spécialisée au domaine de l'agriculture (améliorer une terre → amender, bonifier, fertiliser) et une troisième signification relative à un bien (améliorer un lieu → embellir, réparer, restaurer).

Définition 1 : rendre meilleur, plus satisfaisant, changer, transformer en mieux.

→ perfectionner (améliorer le fonctionnement/l'organisation de qqch) ;

→ corriger, retoucher, réviser, revoir (*améliorer un travail*).

Pour le verbe pronominal, le GR exemplifie l'emploi – *Les relations se sont améliorées*.

Définition 1bis : devenir meilleur.

De même pour le participe passé/adjectif, le GR donne un exemple – *Des hommes améliorés par l'éducation*.

Définition 2 : devenu meilleur.

Figure ci-contre le tableau de synthèse de la signification lexicale du verbe dans sa dimension transitive et pronominal. Nous les regroupons, eu égard à la similarité des définitions.

AMELIORER – S’AMELIORER			
TRAITS NECESSAIRES DE CATEGORISATION (TNC)			
TYPE GRAMMATICAL		ZONE MODALE	ORIENTATION AXIOLOGIQUE
Verbe Transitif – Pronominal		Axiologique	Monovalent – Positif
ENTITES SEMANTIQUES			ENTITES DISCURSIVES
NOYAU		STEREOTYPES	POSSIBLES ARGUMENTATIFS
1	Processus : rendre meilleur, devenir plus satisfaisant	Donc perfectionner... Donc corriger, retoucher, réviser... Donc évoluer en mieux...	A. donc perfectionner un système A. donc corriger un travail S’A. donc bonifier son caractère
2	Résultat : devenu meilleur	Donc perfectionné... Donc corrigé, retouché, révisé... Donc évolué...	A. donc système perfectionné A. donc travail corrigé S’A. donc caractère bonifié

Tableau 4 : Signification lexicale de (s’)améliorer

1.1.2.2 Signification lexicale de amélioration

Comme pour *améliorer*, *amélioration* dans le LL figure une sous-entrée de l’article *meilleur* (cf. Annexe 4). Ce sous-article présente une acception générale du mot que nous conservons pour la construction de la SL et une acception spécialisée au domaine technique du bois (*amélioration des bois*) que nous laissons de côté pour sa référence à un domaine de pratique particulier et pour sa mention limitée au LL. Dans son acception générale, le sous-article ne propose pas de définition, au sens où on n’y trouve pas de périphrase synonymique du défini, mais donne des exemples construits – *L’amélioration des rapports entre les deux états a amené une détente générale*.

Nous nous reportons donc à la définition proposée pour *améliorer*.

Définition : rendre meilleur.

Le GR propose une entrée pour *amélioration* (cf. Annexe 4). L’article se décompose en trois significations : une première signification générale que nous conservons pour la construction de la SL, une deuxième signification spécialisée au domaine de l’agriculture (*amélioration d’un sol* → abonnissement, amendement, bonification) et une troisième signification réservée aux valeurs mobilières et immobilières (*améliorations sur un immeuble* → plus-value ; *amélioration d’une maison* → embellissement, réparation, restauration).

Définition : action de rendre meilleur, de changer en mieux. Fait de devenir meilleur, plus satisfaisant.

→ perfectionnement, progrès (*apporter/faire/opérer/réaliser une amélioration ; amélioration morale/spirituelle/intellectuelle*)

→ correction, retouche, révision (*amélioration d’un travail*).

Les deux dictionnaires décrivent *amélioration* comme un nom féminin. Nous reprenons d'ailleurs cette même terminologie dans nos tableaux de synthèse de la SL. Cette unité lexicale appartient à la zone modale de l'axiologique, puisque le mélioratif *meilleur* participe de sa définition (de son noyau), ce qui, par répercussion, fait de lui un monovalent positif.

Figure ci-dessous le tableau de synthèse de la signification lexicale du substantif.

AMELIORATION(S)		
TRAITS NECESSAIRES DE CATEGORISATION (TNC)		
TYPE GRAMMATICAL	ZONE MODALE	ORIENTATION AXIOLOGIQUE
Nom – Féminin	Axiologique	Monovalent – Positif
ENTITES SEMANTIQUES		ENTITES DISCURSIVES
NOYAU	STEREOTYPES	POSSIBLES ARGUMENTATIFS
Processus : rendre meilleur, plus satisfaisant	Donc perfectionnement, progrès... Donc correction, retouche, révision...	A. donc progrès moral A. donc correction d'un travail

Tableau 5 : Signification lexicale de *amélioration(s)*

1.1.3 Analyse des données

1.1.3.1 Le collocatif est un nom

1.1.3.1.1 Les monovalents positifs

L'application du test de substitution argumentatif – TSA – concerne les collocations : *amélioration de la cohérence*, *amélioration de la pertinence*, *amélioration de la qualité*, *amélioration des performances* et *améliorer l'efficacité*.

Amélioration de la cohérence

Nom : *amélioration de la cohérence* → rendre meilleur, plus satisfaisant → Donc harmonieux, logique, moins contradictoire...

Verbe, processus : *améliorer la cohérence* → rendre meilleur, devenir plus satisfaisant → Donc harmonieux, logique, moins contradictoire...

Amélioration de la pertinence

Nom : *amélioration de la pertinence* → rendre meilleur, plus satisfaisant → Donc adapté...
Donc à-propos, juste...

Verbe, processus : *améliorer la pertinence* → rendre meilleur, devenir plus satisfaisant → Donc adapté... Donc à-propos, juste...

Amélioration de la qualité

Dans la langue générale, *qualité* peut être soit bivalent (employé au sens de *caractéristique*, *valeur*), soit monovalent positif (*bonne qualité*, *qualité positive*). Les occurrences du mot dans le corpus TAL(N) font apparaître un emploi positif quasi systématique.

Ex : *L'intelligibilité et le naturel de la voix synthétique sont les deux principaux critères de qualité de la synthèse.* (AC_0009).

C'est pourquoi nous avons choisi de l'analyser dans cette section et de ne faire figurer que le stéréotype positif.

Nom : *amélioration de la qualité* → rendre meilleur, plus satisfaisant → Donc bonne caractéristique, valeur positive...

Verbe, processus : *améliorer la qualité* → rendre meilleur, devenir plus satisfaisant → Donc bonne caractéristique, valeur positive...

Verbe, résultat : *qualité améliorée* → devenu meilleur → Donc bonne caractéristique, valeur positive...

Amélioration des performances

De même que *qualité*, *performances* (cf. Section 1.1.3.1.2) est soit monovalent positif (employé au sens d'*exploit* ou de *rendement maximal*), soit bivalent (entendu comme *résultat*). Les deux acceptions étant présentes en corpus, nous avons décidé de le faire figurer simultanément dans la section consacrée au monovalents positifs et dans celle consacrée aux bivalents, en y adaptant les enchaînements stéréotypiques respectivement associés.

Nom : *amélioration des performances* → rendre meilleur, plus satisfaisant → Donc exploit, réussite exceptionnelle... Donc rendement maximal, fiabilité élevée...

Verbe, processus : *améliorer les performances* → rendre meilleur, devenir plus satisfaisant → Donc exploit, réussite exceptionnelle... Donc rendement maximal, fiabilité élevée...

Verbe, résultat : *performances améliorées* → devenu meilleur → Donc exploit, réussite exceptionnelle... Donc rendement maximal, fiabilité élevée...

Améliorer l'efficacité

Verbe, processus : *améliorer l'efficacité* → rendre meilleur, devenir plus satisfaisant → Donc produit l'effet attendu... Donc résultat maximum...

L'application du TSA montre que l'adjonction du collocatif à la base *améliorer* et/ou de *amélioration(s)* forme un déploiement argumentatif possible (H32.1). Par conséquent, nous concluons pour chacune de ces collocations à un mécanisme de motivation argumentative (H32.2).

Pour statuer sur l'axiologisation intrinsèque des collocations (H32.3), nous appliquons le test d'enchaînement argumentatif – TEA :

Améliorer la cohérence → Donc progresser

Améliorer la cohérence → *Pourtant progresser

Améliorer la cohérence → *Donc régresser

Améliorer la cohérence → Pourtant régresser

Amélioration de la cohérence → Donc progrès

Amélioration de la cohérence → *Pourtant progrès

Amélioration de la cohérence → *Donc échec

Amélioration de la cohérence → Pourtant échec

Seuls les enchaînements normatif « progrès » et transgressif « échec » sont sémantiquement acceptables, ce qui signifie que le déploiement argumentatif est intrinsèquement axiologisé positivement, en l'absence de tout élément co- ou contextuel.

Au niveau des mécanismes macro-sémantiques d'axiologisation de la signification (H41), la juxtaposition au sein d'une même collocation de deux polarités positives nous permet de statuer sur un mécanisme de renforcement des valeurs axiologiques positives portées par *améliorer* et *amélioration(s)*. Nous attribuons ce mécanisme à toutes les collocations ci-dessus analysées.

Au niveau discursif, l'observation co- et contextuelle des collocations fait apparaître l'existence de deux mécanismes (H42), l'un d'activation des potentiels axiologiques (avec *s'avérer* et deux fois *une source*), l'autre de neutralisation (avec *pouvoir*, *montrer* et *décrire*).

Activation

Ex : *L'intégration de cooccurrences dans les modèles de représentation vectorielle de documents s'est avérée une source d'amélioration de la pertinence des mesures de similarités textuelles* (AC_0049).

Ex : *Dans Salton (69), les sources d'amélioration de la qualité de la recherche d'information en fonction des raffinements sur la fonction d'indexation sont détaillées.* (AR_0013).

Neutralisation

Ex : *Nous pouvons globalement améliorer la cohérence des vecteurs* (AC_0070).

Ex : *Les premiers résultats pour l'évaluation de cette représentation pour une tâche de recherche documentaire montrent une amélioration des performances* (AC_0029).

Ex : *Enfin, nous décrivons, dans la section 4, quelques fonctionnalités qui améliorent l'efficacité ou l'utilisation des analyseurs* (AC_0075).

L'environnement cotextuel de ces collocations apparaît donc pauvre, comme si le mécanisme de renforcement macro-sémantique se suffisait à lui-même, au point de « paralyser » l'axiologisation discursive.

1.1.3.1.2 Les bivalents

L'application du TSA concerne les collocations : *amélioration des performances* et *amélioration des résultats*.

Amélioration des performances

Nom : *amélioration des performances* → rendre meilleur, plus satisfaisant → Donc bon résultat d'une épreuve, d'un test...

Verbe, processus : *améliorer les performances* → rendre meilleur, devenir plus satisfaisant → Donc bon résultat d'une épreuve, d'un test...

Verbe, résultat : *performances améliorées* → devenu meilleur → Donc bon résultat d'une épreuve, d'un test...

Amélioration des résultats

Nom : *amélioration des résultats* → rendre meilleur, plus satisfaisant → Donc effet, conséquence positive...

Verbe, processus : *améliorer les résultats* → rendre meilleur, devenir plus satisfaisant → Donc effet, conséquence positive...

Verbe, résultat : *résultats améliorés* → devenu meilleur → Donc effet, conséquence positive...

Dans la mesure où l'adjonction des collocatifs *performances* et *résultats* à la base *améliorer* et/ou *amélioration* forme un déploiement argumentatif possible (H32.1), les deux collocations correspondent à une motivation argumentative (H32.2).

Pour statuer sur leur axiologisation intrinsèque (H32.3), nous appliquons le TEA :

Améliorer les performances → Donc progresser

Améliorer les performances → *Pourquoi progresser

Améliorer les performances → *Donc régresser

Améliorer les performances → Pourquoi régresser

Amélioration des performances → Donc progrès

Amélioration des performances → *Pourquoi progrès

Amélioration des performances → *Donc échec

Amélioration des performances → Pourquoi échec

De même que pour *amélioration de la cohérence*, seuls les enchaînements normatif « progrès » et transgressif « échec » sont sémantiquement acceptables, ce qui signifie que l'argumentation est intrinsèquement axiologisée positivement. L'orientation axiologique positive des bases *améliorer* et *amélioration* contraint l'instanciation positive des unités bivalentes *performances* et *résultats*. Cette contrainte est inhérente au statut argumentatif des blocs déployés posés comme naturels : processus meilleur → donc résultat meilleurs.

Au niveau macro-sémantique, l'axiologisation positive du bivalent résultant de la monovalence positive de la base, nous statuons sur un mécanisme d'activation des valeurs positives portées par *améliorer* et *amélioration(s)*. Ce même mécanisme régit la relation entre *améliorer-amélioration(s)* et *résultats*.

Au niveau discursif, les informations cotextuelles sont denses et variées. Les extraits ci-dessous sont classés selon les quatre types de mécanismes discursifs recensés, allant du renforcement à l'interversion des potentiels axiologiques (H42).

Renforcement

Ex : *produire une nouvelle **amélioration des résultats*** (AR_0062)

Activation

Ex : *la suppression permet donc une **amélioration des performances*** (AC_0049)

Ex : *une **amélioration des résultats** peut donc être obtenue* (AR_0025)

Affaiblissement

Ex : *il est tout de même possible d'obtenir une **amélioration des performances*** (AC_0056)

Ex : *le couple lui-même apporte donc une légère **amélioration des résultats** de détection de couples qualia* (AR_0060)

Interversion

Ex : *malgré des études prouvant la non-**amélioration des performances** de détection en utilisant des séquences de mots* (AC_0054)

Ex : *la non-**amélioration des résultats** obtenus avec ces procédures de racinisation sur l'anglais* (AR_0016)

Conformément à notre hypothèse H42, ces extraits montrent que les mécanismes de renforcement (avec *nouvelle*) et d'activation (avec *permettre* et *obtenir*) des potentiels axiologiques portés par les collocations sont présents. En revanche, force est de constater que les mécanismes d'affaiblissement (avec *tout de même* et *légère*) et d'interversion (avec *prouver* et deux fois *non*) sont également présents, voire d'un point de vue quantitatif plus représentés que les précédents. Ces emplois apparaissent préférentiellement dans les sections des articles qui discutent des résultats ou concluent dessus, par exemple « Résultats » ou encore « Évaluation et perspectives ». Ce qui laisse penser que l'expérience scientifique, qu'elle se solde positivement ou non, n'en demeure pas moins une avancée méritant d'être communiquée.

1.1.3.1.3 Les neutres

L'application du TSA concerne les collocations : *amélioration de la méthode* et *amélioration des systèmes*.

Amélioration de la méthode

Nom : *amélioration de la méthode* → rendre meilleur, plus satisfaisant → *Donc conduite et expression de la pensée... *Donc procédé d'investigation, moyen...

Nom : *amélioration de la méthode* → rendre meilleur, plus satisfaisant → *Pourtant conduite et expression de la pensée... *Pourtant procédé d'investigation, moyen...

Verbe, processus : *améliorer la méthode* → rendre meilleur, devenir plus satisfaisant → *Donc conduite et expression de la pensée... *Donc procédé d'investigation, moyen...

Verbe, processus : *améliorer la méthode* → rendre meilleur, devenir plus satisfaisant → *Pourtant conduite et expression de la pensée... *Pourtant procédé d'investigation, moyen...

Verbe, résultat : *méthode améliorée* → devenu meilleur → *Donc conduite et expression de la pensée... *Donc procédé d'investigation, moyen...

Verbe, résultat : *méthode améliorée* → devenu meilleur → *Pourtant conduite et expression de la pensée... *Pourtant procédé d'investigation, moyen...

Amélioration des systèmes

Dans le corpus, *système* est employé essentiellement dans son acception spécialisée au domaine de l'informatique.

Nom : *amélioration des systèmes* → rendre meilleur, plus satisfaisant → *Donc traitement matériel et logiciel de données...

Nom : *amélioration des systèmes* → rendre meilleur, plus satisfaisant → *Pourtant traitement matériel et logiciel de données...

Verbe, processus : *améliorer des systèmes* → rendre meilleur, devenir plus satisfaisant → *Donc traitement matériel et logiciel de données...

Verbe, processus : *améliorer des systèmes* → rendre meilleur, devenir plus satisfaisant → *Pourtant traitement matériel et logiciel de données...

Dans ces deux cas, l'application du TSA montre que l'adjonction du collocatif à la base ne forme pas avec les stéréotypes proposés ou envisageables de *améliorer* et/ou de *amélioration(s)* un bloc d'argumentation interne, ni un déploiement argumentatif possible (H33.1). Ces collocations ne peuvent être sémantiquement prédictibles ; elles n'ont de réalité que discursive. Par conséquent, ces collocations sont qualifiées d'arbitraires (H33.2).

Nous rappelons que par l'absence de lien argumentatif, la collocation arbitraire ne devrait pas être intrinsèquement axiologisée. Toutefois, sans remettre en cause la dimension évaluative de la langue générale, nous pensons que, bien que neutralisée sémantiquement, la base conserve son potentiel axiologique (H33.3). Nous appliquons donc le TEA :

Améliorer la méthode → Donc progresser

Améliorer la méthode → *Pourtant progresser

Améliorer la méthode → *Donc régresser

Améliorer la méthode → Pourtant régresser

Amélioration de la méthode → Donc progrès

Amélioration de la méthode → *Pourtant progrès

Amélioration de la méthode → *Donc échec

Amélioration de la méthode → Pourtant échec

De même que pour les monovalents positifs et les bivalents, seuls les enchaînements normatif « progrès » et transgressif « échec » sont sémantiquement acceptables, ce qui signifie que la collocation est intrinsèquement axiologisée positivement, sans être intrinsèquement argumentative, mais tout en conservant son potentiel axiologique. L'orientation axiologique positive des bases *améliorer* et *amélioration* semble contraindre l'instanciation positive de l'ensemble de la collocation, sans que nous ayons la preuve qu'elle « contamine » la neutralité du collocatif.

Au niveau discursif, l'observation co- et contextuelle des collocations fait apparaître l'existence de deux mécanismes (H42), l'un de renforcement des potentiels axiologiques (avec *réduire les coûts*), l'autre d'activation (avec *réussir*).

Renforcement

Ex : *Les coûts de mise en place peuvent être réduits [...] par l'adoption et l'amélioration de méthodes de génie logiciel adaptées aux applications de génération 9.* (AR_0026)

Activation

Ex : *nous avons réussi à améliorer le système en six mois* (AC_0067)

1.1.3.2 Le collocatif est un adjectif

1.1.3.2.1 Les monovalents positifs

L'application du TSA concerne les collocations : *amélioration sensible*, *amélioration significative* et *amélioration substantielle*.

Amélioration sensible

Nom : *amélioration sensible* → rendre meilleur, plus satisfaisant → Donc grand, intense, important, manifeste...

Verbe, processus : *améliorer sensiblement* → rendre meilleur, devenir plus satisfaisant → Donc changer d'une manière appréciable, importante, notable...

Amélioration significative

Nom : *amélioration significative* → rendre meilleur, plus satisfaisant → Donc important, marquant...

Verbe, processus : *améliorer significativement* → rendre meilleur, devenir plus satisfaisant → Donc changer d'une manière importante, marquante...

Amélioration substantielle

Nom : *amélioration significative* → rendre meilleur, plus satisfaisant → Donc fondamental, essentiel...

Verbe, processus : *améliorer significativement* → rendre meilleur, devenir plus satisfaisant → Donc changer d'une manière fondamentale, essentielle...

L'application du TSA montre que l'adjonction du collocatif à la base *améliorer* et/ou de *amélioration(s)* un déploiement argumentatif possible (H32.1). Par conséquent, ces trois collocations correspondent à un mécanisme de motivation argumentative (H32.2).

Pour statuer sur leur axiologisation intrinsèque (H32.3), nous appliquons le TEA :

Améliorer sensiblement → Donc progresser

Améliorer sensiblement → *Pourquoi progresser

Améliorer sensiblement → *Donc régresser

Améliorer sensiblement → Pourquoi régresser

Amélioration sensible → Donc progrès

Amélioration sensible → *Pourquoi progrès

Amélioration sensible → *Donc échec

Amélioration sensible → Pourquoi échec

Là encore, seuls les enchaînements normatif « progrès » et transgressif « échec » sont sémantiquement acceptables, ce qui signifie que l'argumentation est intrinsèquement axiologisée positivement.

Au niveau des mécanismes macro-sémantiques d'axiologisation de la signification, la polarité positive de chacun des collocatifs *sensible*, *significative*, et *substantielle* s'adjoint à celle des bases *améliorer* et *amélioration*, également positives, vis-à-vis desquelles ils fonctionnent comme des intensifieurs. De même que pour *amélioration de la cohérence*, l'association de deux polarités positives produit un effet de renforcement du degré d'axiologisation (H41).

Au niveau discursif, l'étude co- et contextuelle des collocations fait apparaître l'existence de trois mécanismes (H42) : l'activation (avec *apporter*), la neutralisation (avec *présenter*) et l'interversion (avec *ne pas fournir*), les deux premiers étant les plus utilisés.

Activation

Ex : *l'utilisation des CS apporte une **amélioration sensible*** (AC_0085)

Neutralisation

Ex : *ces résultats présentent une **amélioration significative*** (AR_0018)

Interversion

Ex : *la racinisation ne fournissait pas d'**amélioration substantielle*** (AR_0016)

1.1.3.2.2 Le monovalent négatif

L'application du TSA concerne la collocation *légère amélioration*.

Légère amélioration

Nom : *légère amélioration* → rendre meilleur, plus satisfaisant → *Donc peu important...

Nom : *légère amélioration* → rendre meilleur, plus satisfaisant → Pourquoi peu important...

Dans sa forme normative, le déploiement argumentatif est asémantique. En revanche, dans sa forme transgressive, il recouvre une valeur sémantique correcte. Aussi, l'application du TSA montre que l'adjonction du collocatif *légère* à la base *amélioration* forme un déploiement argumentatif possible (H32.1). Il s'agit d'un mécanisme de motivation argumentative (H32.2).

Pour statuer sur son axiologisation intrinsèque (H32.3), nous appliquons le TEA :

Légère amélioration → ?Donc progrès

Légère amélioration → *Pourtant progrès

Légère amélioration → *Donc échec

Légère amélioration → ?Pourtant échec

À l'inverse des collocations étudiées auparavant, seuls les enchaînements normatif « progrès » et transgressif « échec » sont sémantiquement acceptables, mais avec une certaine « réserve ». De fait, *légère* est un monovalent négatif, par conséquent, son adjonction à *amélioration* confère à la collocation une valeur axiologique intrinsèque positive amoindrie. Ce paradigme collocationnel permet de formuler une nouvelle hypothèse, en lien avec la polarité de la base, qui serait « plus forte » que celle du collocatif, au sens où elle est susceptible d'imposer son orientation axiologique à l'ensemble de la collocation. Nous vérifierons donc la validité de ce processus dans les cas de polarités opposées au sein d'une même collocation.

Au niveau du mécanisme macro-sémantique d'axiologisation de la signification, la polarité négative du collocatif *légère* s'adjoint à la polarité positive de la base *amélioration*, vis-à-vis de laquelle il fonctionne comme un atténuateur. Nous statuons donc sur un mécanisme d'affaiblissement de la polarité positive portée par la base.

Au niveau discursif, l'observation co- et contextuelle de la collocation fait apparaître l'existence de deux mécanismes (H42), l'un d'activation du potentiel axiologique positif de la collocation (avec *apporter*), l'autre de neutralisation (avec *présenter*).

Activation

Ex : *le couple lui-même apporte donc une légère amélioration* (AR_0060)

Neutralisation

Ex : *On constate une légère amélioration avec le recours aux pondérations* (AR_0062)

1.2 Performance(s)

1.2.1 Dépouillement des données

1.2.1.1 Fréquences et nettoyage

Le corpus TAL(N) comporte 112 occurrences de *performance* au singulier et 321 occurrences du mot au pluriel.

Selon le critère statistique, nous avons éliminé les candidats trop peu fréquents (apax, cooccurents de fréquence deux...).

Ex : *performance appliquée* (1), *performance de calcul* (1), *performances honorables* (1), *performances élevées* (2), *performances de détection* (3, AC_0054).

Selon le critère de pertinence, nous laissons de côté toutes les collocations grammaticales et autres constructions dont le collocatif ne présente pas réellement de contenu sémantique.

Ex : *performance sur*, *performances de/des/du*.

Aucune lexie apparentée au critère lexical n'a été recueillie, de même qu'aucun terme complexe, construit sur *performance(s)* et entrant dans la catégorie du critère lexicographique, n'a été relevé dans le corpus. Toutefois, nous avons retiré la collocation *variation(s) de performance(s)*, *performance* étant employé au sens linguistique du terme, sans pour autant représenter un terme complexe.

1.2.1.2 Liste des collocations lexicales

Cette sous-section présente les quatre tableaux récapitulatifs des collocations lexicales construites sur *performance(s)*. Les deux premiers tableaux correspondent au contexte droit et les deux derniers, au contexte gauche.

COLLOCATIONS LEXICALES	FREQ.	AUTRES FORMES	FREQ.	TOTAL
Performance de l'algorithme	1	-P. des algorithmes -Performances de l'algorithme -Performances des algorithmes	1 1 2	5
Performance des systèmes	3	-P. de leur système -P. de notre système -P. du système -P. d'un système -Performances des systèmes -Performances [...] des systèmes -Performances d'un système -Performances du système -Performances de [...] système(s)	1 3 3 1 18 6 7 9 9	57
Performance des méthodes	-	-P. de ces méthodes -P. [...] de cette méthode -P. de nos méthodes -Performances des méthodes -Performances de [...] méthode(s) -Performances [...] de la méthode	1 1 1 3 6 1	13
Performance des modules	-	-P. des [...] modules -Performances des modules -Performances du module	1 1 2	4
Performance du modèle	-	-P. [...] du modèle -Performances du modèle -Performances [...] du modèle	1 2 2	5
Performance globale	3	-Performances globales	6	9
Performance intéressante	1	-Performances intéressantes -Performances [...] intéressantes	2 1	4
Performance moyenne	7	-Performances moyennes	4	11
Performance réelle	1	-Performances réelles	3	4
Performance supérieure	1	-Performances supérieures -Performances [...] supérieures	2 4	7

Tableau 6 : Collocations lexicales contexte droit de performance(s)

COLLOCATIONS LEXICALES	FREQ.	AUTRES FORMES	FREQ.	TOTAL
Performances d'analyse	1	-P. de l'analyse -P. de notre analyse	1 1	3
Performances meilleures	1	-P. [...] meilleures -meilleures P.	4 8	13
Performances de l'analyseur	2	-P. de notre analyseur	1	3
Performances de l'étiquetage	2	-P. d'étiquetage	2	4

Tableau 7 : Collocations lexicales contexte droit de performances

COLLOCATIONS LEXICALES	FREQ.	AUTRES FORMES	FREQ.	TOTAL
Baisse de performance	5	-Baisse de performances -Baisse des performances	1 2	8
Bonne performance	2	-Bonnes performances -P. moins bonnes -P. [...] bonnes	5 1 2	10
Gain de performance	1	-Gain en P. -Gain en performances	2 1	4
Mesure de performance	8	-Mesure(s) de performances -Mesurer + performances	7 5	20

Tableau 8 : Collocations lexicales contexte gauche performance(s)

COLLOCATIONS LEXICALES	FREQ.	AUTRES FORMES	FREQ.	TOTAL
Comparaison des performances	3	-	-	3
Dégradation des performances	4	-Dégrader + P.	3	7
Diminution des performances	2	-Diminuer + performances	1	3

Tableau 9 : Collocations lexicales contexte gauche performances

1.2.2 Signification lexicale de performance

L'unité lexicale *performance* bénéficie d'une entrée à part entière dans le LL, sous laquelle figure également l'adjectif *performant(e)* et le nom *contre-performance*, que nous laissons de côté (*cf.* Annexe 5). L'article sépare les définitions consacrées au nom singulier de celles consacrées au pluriel.

Au singulier, on compte cinq définitions dont l'une consacrée à une lexie restreinte au domaine de la psychologie – *test de performance*. Parmi les quatre définitions restantes, deux d'entre elles sont spécialisées, respectivement associées au domaine sportif (*performance d'un athlète, d'une équipe*) et au domaine linguistique (par opposition à *compétence*)³. Les deux autres sont générales ; nous les conservons pour la construction de la SL. La première SL désigne *performance* au sens d'*exploit*. En revanche, la seconde attribue au mot *performance* le statut de bivalent – résultat obtenu au cours d'une épreuve, d'un test. Sur les conseils des experts du domaine et conformément aux emplois du mot en co- et contexte, nous choisissons de conserver cette SL, en lui réservant toutefois un tableau à part, dans la mesure où l'orientation axiologique diffère des autres SL.

Au pluriel, l'article comporte deux définitions, l'une spécialisée au domaine de l'aéronautique – Chiffres et courbes qui caractérisent les possibilités de vol d'un avion –, que nous laissons de côté ; l'autre, générale, désigne les qualités (accélération, vitesse maximale, décélération...) qui caractérisent la valeur d'une voiture, que nous gardons en élargissant sa portée conformément aux propositions du GR (mentionnée avec les informations relatives au GR).

Définition 1 : exploit.

Définition 2 : résultat obtenu au cours d'une épreuve, d'un test.

³ Le dépouillement des données de notre corpus montre que *performance*, au sens linguistique du terme, ne constitue pas une base propre à générer des collocations lexicales. Aussi, nous laissons sa définition de côté pour la construction de la SL.

Le GR propose une entrée à part entière pour *performance* (cf. Annexe 5). L'article se décompose également en cinq significations. Sur ces cinq définitions, trois sont spécialisées au domaine sportif (*performance d'un cheval* → bonne ou mauvaise performance ; *performances des cyclistes* → performances homologuées), au domaine de la psychologie (*besoin de performance de l'individu*) et au domaine linguistique (par opposition à *compétence*) ; nous les laissons toutes trois de côté. Enfin, deux sont générales, l'une désigne *performance* au sens de exploit, réussite remarquable dans tout autre domaine, l'autre est théoriquement réservée à l'acception technique du mot – rendement maximal, fiabilité élevée (d'une machine, de l'homme) (performance d'un ordinateur) ; indication quantitative sur les caractéristiques d'un véhicule, d'un système mécanique (vitesse, autonomie, rayon d'action...).

Définition 1 : exploit, réussite remarquable dans tout autre domaine.

→ accomplir, réaliser (*performance homologuée*)

→ record, succès (*belle performance* ; *performance érotique*)

Définition 2 : rendement maximal, fiabilité élevée...

Cette seconde définition reprend les mêmes caractéristiques que la dernière définition proposée pour le LL, en les élargissant à d'autres entités que les voitures, nous choisissons de les regrouper.

Les deux dictionnaires décrivent *performance* comme un nom féminin. Cette unité lexicale appartient à la zone modale de l'axiologique soit en qualité de monovalent positif, soit en qualité de bivalent, comme l'attestent les synonymes proposés.

Figure ci-dessous les tableaux de synthèse de la signification lexicale du nom.

PERFORMANCE(S)			
TRAITS NECESSAIRES DE CATEGORISATION (TNC)			
TYPE GRAMMATICAL		ZONE MODALE	ORIENTATION AXIOLOGIQUE
Nom – Féminin		Axiologique	Monovalent – Positif
ENTITES SEMANTIQUES			ENTITES DISCURSIVES
NOYAU		STEREOTYPES	POSSIBLES ARGUMENTATIFS
1	Exploit, réussite exceptionnelle	Donc accomplissement, contribution, réalisation... Donc prouesse, record, succès...	P. donc contribution homologuée P. donc prouesse amoureuse
2	Rendement maximal, fiabilité élevée d'une machine, d'un objet ou d'un matériau	Donc amélioration, progression... Donc efficacité, optimisation...	P. donc amélioration des matériels P. donc efficacité d'un ordinateur

Tableau 10 : Signification lexicale 1 de performance(s)

PERFORMANCE(S)		
TRAITS NECESSAIRES DE CATEGORISATION (TNC)		
TYPE GRAMMATICAL	ZONE MODALE	ORIENTATION AXIOLOGIQUE
Nom – Féminin	Axiologique	Bivalent
ENTITES SEMANTIQUES		ENTITES DISCURSIVES
NOYAU	STEREOTYPES	POSSIBLES ARGUMENTATIFS
Résultat : résultat obtenu au cours d'une épreuve, d'un test	Donc bon ou mauvais... Donc conséquence, effet, suite... Donc analyse, comparaison, interprétation, quantification... Donc discussion, diffusion...	P. donc conséquence espérée P. donc effet fâcheux P. donc interprétation décisive P. donc discussion intéressante

Tableau 11 : Signification lexicale 2 de performance(s)

L'aspect dual, monovalent positif-bivalent, de la signification lexicale de performance nous oblige à observer chaque collocation en co- et contexte de façon à pouvoir statuer sur l'emploi mobilisé, en vue d'une application cohérente du TSA. À l'issue de cette recherche, dans la plupart des cas, il se révèle qu'une même collocation sélectionne *performance* soit dans l'acception de *rendement maximal* (SL 1 – noyau 2), soit dans l'acception de *résultat* (SL 2). C'est pourquoi l'analyse argumentative reprend pour la majorité des collocations ces deux SL. Toutefois, dans les cas où une collocation n'admet qu'une seule SL, nous le mentionnons et le justifions par un exemple.

1.2.3 Analyse des données

1.2.3.1 Le collocatif est un nom

1.2.3.1.1 Le monovalent positif

L'application du TSA concerne la collocation *gain de performance*⁴.

Gain de performance

Les emplois de cette collocation montrent que la performance dont il est question réfère à une technique logicielle, du type *logiciel d'étiquetage stochastique* (AC_0116). Aussi, nous ne retenons que la SL 1 – noyau 2.

⁴ Nous n'analysons pas *amélioration des performances*, puisque *amélioration* fait partie des six bases collocationnelles retenues ; par conséquent, cette collocation a déjà été analysée. Pour la même raison, nous analysons *évaluation de la performance* non pas dans cette section mais dans celle consacrée à l'unité lexicale *évaluation*.

SL 1 – noyau 2 : *gain de performance* → rendement maximal, fiabilité élevée → Donc profit, succès, réussite, progrès...

L'application du TSA montre que l'adjonction du collocatif à la base forme avec les stéréotypes proposés ou envisageables de *performance(s)* un bloc d'argumentation interne (H31.1). Par conséquent, nous concluons à la concrétisation d'un PA dans le discours (H31.2). Cette collocation est argumentative ; elle forme un bloc d'argumentation externe.

Pour statuer sur son axiologisation intrinsèque (H31.3), nous appliquons le TEA :

Gain de performance → Donc progrès

Gain de performance → *Pourtant progrès

Gain de performance → *Donc échec

Gain de performance → Pourtant échec

Les enchaînements normatif « progrès » et transgressif « échec » sont sémantiquement acceptables, ce qui signifie que le PA est intrinsèquement axiologisé positivement, en l'absence de tout élément co- ou contextuel.

Au niveau des mécanismes macro-sémantiques d'axiologisation de la signification (H41), comme pour *amélioration de la cohérence*, la juxtaposition au sein d'une même collocation de deux polarités positives nous permet de statuer sur un mécanisme de renforcement des valeurs axiologiques positives portées par *performance(s)*.

Au niveau discursif, l'observation co- et contextuelle des collocations fait apparaître l'existence de deux mécanismes (H42), l'un de neutralisation des potentiels axiologiques (avec le bivalent *évaluer*), l'autre d'affaiblissement (avec *faible*).

Neutralisation

Ex : *nous avons évalué le **gain en performances** obtenu pour les deux types de vocabulaires étudiés* (AC_0080)

Affaiblissement

Ex : *du fait du **faible gain en performance** observé dans nos expériences* (AC_0049)

1.2.3.1.2 Les monovalents négatifs

L'application du TSA concerne les collocations : *baisse de performance*, *dégradation des performances* et *diminution des performances*.

Baisse de performance

SL 1 – noyau 2 : *baisse de performance* → rendement maximal, fiabilité élevée → *Donc diminution de force, réduction d'intensité, descente du niveau...

SL 1 – noyau 2 : *baisse de performance* → rendement maximal, fiabilité élevée → Pourtant diminution de force, réduction d'intensité, descente du niveau...

SL 2 : *baisse de performance* → résultat d'une épreuve, d'un test → Donc diminution de force, réduction d'intensité, descente du niveau...

Dégradation des performances

SL 1 – noyau 2 : *dégradation des performances* → rendement maximal, fiabilité élevée → *Donc affaiblissement, diminution, perte de valeur...

SL 1 – noyau 2 : *dégradation des performances* → rendement maximal, fiabilité élevée → Pourtant affaiblissement, diminution, perte de valeur...

SL 2 : *dégradation des performances* → résultat d'une épreuve, d'un test → Donc affaiblissement, diminution, perte de valeur...

Diminution des performances

SL 1 – noyau 2 : *diminution des performances* → rendement maximal, fiabilité élevée → *Donc baisse, décroissance, réduction...

SL 1 – noyau 2 : *diminution des performances* → rendement maximal, fiabilité élevée → Pourtant baisse, décroissance, réduction...

SL 2 : *diminution des performances* → résultat d'une épreuve, d'un test → Donc baisse, décroissance, réduction...

Concernant l'emploi de *performance(s)* en qualité de monovalent positif, nous constatons que pour chacune des trois collocations le bloc d'argumentation est asémantique dans sa forme normative. Mais, en déployant le bloc d'argumentation dans sa forme transgressive, il recouvre une valeur sémantique correcte. Aussi, l'application du TSA montre que l'adjonction du collocatif à la base forme avec les stéréotypes proposés ou envisageables de *performance(s)* un bloc d'argumentation interne (H31.1). Par conséquent, nous concluons à la concrétisation d'un PA dans le discours (H31.2). Ces collocations sont argumentatives ; elles forment des blocs d'argumentations externes.

Pour statuer sur leur axiologisation intrinsèque (H31.3), nous appliquons le TEA :

Baisse de performance → *Donc progrès

Baisse de performance → Pourtant progrès

Baisse de performance → Donc échec

Baisse de performance → *Pourtant échec

Seuls les enchaînements transgressif « progrès » et normatif « échec » sont sémantiquement acceptables. De fait, *baisse*, *dégradation* et *diminution* sont des monovalents négatifs ; par conséquent, leur adjonction à *performance* confère à la collocation une valeur axiologique intrinsèque négative. L'hypothèse posée concernant la polarité de la base qui serait plus forte que celle du collocatif est donc invalidée.

Au niveau des mécanismes macro-sémantiques d'axiologisation de la signification, nous statuons sur un mécanisme d'interversion des valeurs portées positives portées par la base.

L'environnement discursif de la collocation fait apparaître un seul mécanisme discursif de neutralisation porté par des unités comme *entraîner* et *être un facteur de*.

Neutralisation

Ex : *Traiter les mots sans tenir compte de leur position par rapport au mot à désambiguïser entraîne également une **baisse des performances*** (AC_0079).

Ex : *Enfin, l'absence de retour de pertinence dans le système MEDLARS est aussi un facteur de **dégradation des performances*** (AR_0013).

Ex : *Il apparaît donc que, si un filtrage syntaxique peut être bénéfique, la mise en œuvre de contraintes de sélection trop fortes entraîne une **diminution des performances*** (AC_0049).

Dans l'acception bivalente de *performance(s)*, l'enchaînement argumentatif en *donc* formé par l'adjonction du collocatif à la base forme immédiatement un bloc d'argumentation interne (H31.1). Par conséquent, nous concluons à la concrétisation de PA dans le discours (H31.2). Ces collocations sont argumentatives ; elles forment des blocs d'argumentations externes.

Pour statuer sur leur axiologisation intrinsèque (H31.3), nous appliquons le TEA :

Baisse de performance → *Donc progrès

Baisse de performance → Pourtant progrès

Baisse de performance → Donc échec

Baisse de performance → *Pourtant échec

À l'inverse de l'analyse précédente, seuls les enchaînements transgressif « progrès » et normatif « échec » sont sémantiquement acceptables. De fait, *baisse*, *dégradation* et *diminution* étant monovalents négatifs, leur adjonction à *performance(s)* confère aux collocations une valeur négative (H31.3). Une *baisse*, une *dégradation* ou encore une *diminution* « contaminent négativement » la bivalence de *performance*. Ce paradigme collocationnel laisse penser que la bivalence de la base est contaminée lorsque le collocatif est axiologiquement marqué. Nous vérifierons là encore la validité de ce processus dans les cas de polarités positive ou négative adjointes à une base bivalente (cf. Section 3).

Eu égard au phénomène de contamination décrit ci-dessus, au niveau macro-sémantique, nous statuons sur un mécanisme d'activation négative du potentiel axiologique bivalent de la base *performance(s)*.

Au niveau discursif, l'étude co- et contextuelle des collocations fait apparaître l'existence de deux mécanismes (H42), l'un de renforcement (avec *mauvais* suivi de *gêner* et *privilégier trop la réduction*), l'autre de neutralisation (avec *être cause de*).

Renforcement

Ex : *ce mauvais étiquetage peut en effet gêner et **diminuer les performances*** (AC_0037)

Ex : *dans nos expériences, le filtrage des cooccurrences par les groupes syntaxiques semble constituer un compromis efficace alors que la sélection des cooccurrences par les relations syntaxiques privilégie trop la réduction d'information et mène donc à une **dégradation des performances*** (AC_0049)

Neutralisation

Ex : *les textes 8 et 3 sont cause de **baisse de performances*** (AC_0016)

1.2.3.1.3 Le bivalent

L'application du TSA concerne *performance du modèle*. La bivalence de l'unité lexicale *modèle* tient à la présence dans les dictionnaires de définition à alternances « *formalisé ou non* », « *scientifiquement testé ou non* ». Aussi, l'instanciation axiologique positive ou négative est possible.

Performance du modèle

SL 1 – noyau 2 : *performance du modèle* → rendement maximal, fiabilité élevée → Donc processus, système formalisé, scientifiquement testé...

L'application du TSA pour la SL 2 de *performance du modèle* pourrait poser question dans la mesure où la base et le collocatif sont tous deux bivalents. Or les éléments de signification associés à chacune des unités nous permettent de trouver une solution naturelle. Étant donné que l'obtention de résultat nécessite préalablement la mise en place d'un processus à tester, nous instancions logiquement la polarité positive de *modèle* pour cet enchaînement argumentatif.

SL 2 : *performance du modèle* → résultat d'une épreuve, d'un test → Donc processus, système formalisé, scientifiquement testé...

Pour les deux SL, l'application du TSA montre que l'adjonction du collocatif *modèle* à la base *performance(s)* forme un déploiement argumentatif possible (H32.1). Par conséquent, nous concluons à un mécanisme de motivation argumentatif (H32.2).

L'axiologisation intrinsèque (H32.3) opère différemment selon la SL de *performance(s)* qui est mobilisée.

Nous appliquons le TEA à la SL – noyau 2 :

Performance du modèle → Donc progrès

Performance du modèle → *Pourtant progrès

Performance du modèle → *Donc échec

Performance du modèle → Pourtant échec

Seuls les enchaînements normatif « progrès » et transgressif « échec » sont sémantiquement acceptables. De fait, *performance* est monovalent positif, ce qui contraint l'instanciation positive de l'unité bivalente *modèle*. La polarité positive de la base contamine la bivalence du collocatif. La collocation est donc intrinsèquement axiologisée positivement.

Nous appliquons le TEA à la SL 2 :

Performance du modèle → Donc progrès

Performance du modèle → Pourtant progrès

Performance du modèle → Donc échec

Performance du modèle → Pourtant échec

Dans ce cas, la base et le collocatif sont tous deux bivalents, c'est pourquoi tous les enchaînements, qu'ils soient normatifs ou transgressifs, sont sémantiquement acceptables. La collocation est donc intrinsèquement axiologisée, sans que l'un ou l'autre des pôles positif ou négatif soient activés.

Au niveau macro-sémantique, dans le cas de la SL 1 – noyau 2, l'axiologisation positive du bivalent résultant de la monovalence positive de la base, nous statuons sur un mécanisme d'activation de l'axiologisation de la signification.

Dans le cas de la SL 2, nous statuons sur un mécanisme de neutralisation des potentiels axiologiques. Seul un mécanisme discursif peut rompre la bivalence de la collocation.

Au niveau discursif, dans le cas de la SL 1 – noyau 2, on constate un mécanisme de neutralisation (avec *évaluer*) ; dans le cas de la SL 2, un mécanisme d'activation négative (avec être dépassé).

Neutralisation

Ex : *évaluer de façon expérimentale les performances du modèle DOP* (AC_0034)

Activation négative

Ex : *les performances du modèle cache ont été dépassées* (AC_0054)

1.2.3.1.4 Les neutres

L'application du TSA concerne les six collocations suivantes : *performance de l'algorithme*, *performance des systèmes*, *performance des méthodes*, *performance des modules*, *mesure de performance* et *comparaison des performances*.

Performance de l'algorithme

SL 1 – noyau 2 : *performance de l'algorithme* → rendement maximal, fiabilité élevée → Donc calcul, enchaînement des actions nécessaires à l'accomplissement d'une tâche...

SL 2 : *performance de l'algorithme* → résultat d'une épreuve, d'un test → Donc calcul, enchaînement des actions nécessaires à l'accomplissement d'une tâche...

Performance des systèmes

Lorsque *performance* est employé au sens de la SL 1 – noyau 2, *système*, principalement employé au pluriel, pourrait se substituer à *logiciel* :

Ex : *système de recherche* (AR_0018), *système d'étiquetage* (AC_0023), *système de résumé automatique* (AC_0036), *système de transcription automatique* (AR_0034).

À l'inverse, quand *performance* est employé au sens de la SL 2, c'est lui qui apparaît au pluriel.

Ex : *évaluer les performances* (AC_0117), *augmenter les performances* (AR_0060), *mesurer les performances* (AC_0117).

SL 1 – noyau 2 : *performance des systèmes* → rendement maximal, fiabilité élevée → Donc traitements matériels et logiciels de données...

SL 2 : *performance des systèmes* → résultat d'une épreuve, d'un test → Donc traitements matériels et logiciels de données...

Performance des méthodes

Lorsque *performance* est employé au sens de la SL 1 – noyau 2, *méthodes*, principalement employé au pluriel, est employé au sens de *procédé technique* :

Ex : *méthode algorithmique* (AC_0001), *méthode d'alignement* (AC_0008), *méthode d'identification* (AC_0006), *méthode d'apprentissage* (AC_0011).

De même que pour *performance des systèmes*, pour *performance des méthodes* lorsque *performance* est employé au sens de la SL 2, c'est lui qui apparaît au pluriel.

Ex : *analyser les performances* (AR_0053), *maximiser les performances* (AC_0080), *calculer les performances* (AR_0012).

SL 1 – noyau 2 : *performance des méthodes* → rendement maximal, fiabilité élevée → Donc conduite et expression de la pensée... Donc procédé d'investigation, moyens...

SL 2 : *performance des systèmes* → résultat d'une épreuve, d'un test → Donc conduite et expression de la pensée... Donc procédé d'investigation, moyens...

Performance des modules

Un *module* représente un sous-ensemble d'un programme.

Ex : *module de stylisation* (AC_0033), *module de compréhension* (AR_0054).

SL 1 – noyau 2 : *performance des modules* → rendement maximal, fiabilité élevée → Donc calcul, programme destiné à accomplir des tâches prédéfinies...

SL 2 : *performance des modules* → résultat d'une épreuve, d'un test → Donc calcul, programme destiné à accomplir des tâches prédéfinies...

Mesure de performance

Dans le cas de cette collocation, l'étude co- et contextuelle nous apprend que *performance* désigne le plus souvent les résultats des logiciels comme *les étiqueteurs morphosyntaxiques du français* (AC_0023). Aussi, nous ne retenons que la SL 2.

SL 2 : *mesure de performance* → résultat d'une épreuve, d'un test → Donc évaluation, comparaison de la grandeur, de la quantité...

Comparaison des performances

De même que pour *mesure de performance*, l'étude co- et contextuelle de *comparaison des performances* nous pousse à ne conserver que la SL 2 :

Ex : *L'évaluation de cette capacité est vue ici comme la comparaison de la réponse du système avec une réponse attendue par l'évaluateur.* (AR_0053)

SL 2 : *comparaison des performances* → résultat d'une épreuve, d'un test → Donc analyse, jugement, rapprochement...

Pour les deux SL, l'application du TSA montre que l'adjonction du collocatif à la base forme *performance(s)* forme un déploiement argumentatif possible (H32.1). Par conséquent, nous concluons à un mécanisme de motivation argumentative (H32.2).

Là encore, l'axiologisation intrinsèque (H32.3) opère différemment selon la SL de *performance(s)* qui est mobilisée.

Nous appliquons le TEA à la SL – noyau 2 :

Performance de l'algorithme → Donc progrès

Performance de l'algorithme → *Pourtant progrès

Performance de l'algorithme → *Donc échec

Performance de l'algorithme → Pourtant échec

De même que pour les monovalents positifs et les bivalents, seuls les enchaînements normatif « progrès » et transgressif « échec » sont sémantiquement acceptables, ce qui signifie que la collocation est intrinsèquement axiologisée positivement. L'orientation axiologique positive de la base *performance* contraint l'instanciation positive de l'ensemble de la collocation, sans que nous ayons la preuve qu'elle « contamine » la neutralité des collocatifs.

Nous appliquons le TEA à la SL 2 :

Performance de l'algorithme → Donc progrès

Performance de l'algorithme → Pourtant progrès

Performance de l'algorithme → Donc échec

Performance de l'algorithme → Pourtant échec

Dans ce cas, la base est bivalente et le collocatif est neutre. Là encore, tous les enchaînements, qu'ils soient normatifs ou transgressifs, sont sémantiquement acceptables. La collocation est donc porteuse d'un potentiel axiologique intrinsèque, sans que l'un ou l'autre des pôles positif ou négatif ne soient activés.

Au niveau macro-sémantique, dans le cas de la SL 1 – noyau 2, l'axiologisation positive de la collocation résultant de la monovalence positive de la base, nous statuons sur un mécanisme d'activation de l'axiologisation de la signification.

Dans le cas de la SL 2, nous statuons sur un mécanisme de neutralisation des potentiels axiologiques. Seul un mécanisme discursif peut rompre la bivalence de la collocation.

Au niveau discursif, les informations cotextuelles montrent une grande richesse des mécanismes discursifs, quelle que soit la signification concernée, comme en témoignent les extraits ci-dessous classés par collocation et allant du renforcement (avec *apprécier les très bonnes, permettre de maximiser*) à l'interversion (avec *détériorer de façon significative*), en passant par l'activation (avec *améliorer, utiliser* et *offrir*), la neutralisation (avec *comparaison, évaluer, remarquer*) et l'affaiblissement (avec *ralentir*).

Renforcement

Ex : *pour pouvoir apprécier les très bonnes **performances du système*** (AC_0078)

Ex : *jusqu'à présent, nous avons cherché les vocabulaires qui permettaient de maximiser les **performances des méthodes** indépendamment les unes des autres* (AC_0080).

Activation

Ex : *ces informations complémentaires influencent légèrement les résultats en améliorant les **performances des algorithmes*** (AC_0101)

Ex : *le paradigme offre donc une analyse qualitative et diagnostique automatique des **performances du module*** (AR_0054)

Ex : *utiliser comme mesure étalon (gold standard) la **mesure de performance** d'un opérateur humain* (AR_0054)

Neutralisation

Ex : *comparaison des **performances des algorithmes*** (AC_0101)

Ex : *la première expérience évalue les **performances du système** PROMÉTHÉE* (AC_0115)

Ex : *nous pouvons remarquer que les **performances des méthodes** varient significativement d'un thème à l'autre* (AC_0054)

Ex : *afin de permettre la **comparaison des performances** sur des textes différents* (AR_0053)

Affaiblissement

Ex : *la traduction des requêtes ralentit les **performances du système*** (AR_0016)

Interversion

Ex : *le filtrage du deuxième type semble quant à lui trop restrictif et détériore de façon significative les **performances du système*** (AC_0049)

1.2.3.2 Le collocatif est un adjectif

1.2.3.2.1 Les monovalents positifs

L'application du TSA concerne les collocations *performance intéressante, performance supérieure, performances meilleures* et *bonne performance*.

Performance intéressante

SL 1 – noyau 2 : *performance intéressante* → rendement maximal, fiabilité élevée → Donc captivant... Donc important... Donc avantageux, valable...

SL 2 : *performance intéressante* → résultat d'une épreuve, d'un test → Donc captivant... Donc important... Donc avantageux, valable...

Performance supérieure

SL 1 – noyau 2 : *performance supérieure* → rendement maximal, fiabilité élevée → Donc plus grand, plus élevé... Donc meilleur...

SL 2 : *performance supérieure* → résultat d'une épreuve, d'un test → Donc plus grand, plus élevé... Donc meilleur...

Performances meilleures

SL 1 – noyau 2 : *performances meilleures* → rendement maximal, fiabilité élevée → Donc de qualité supérieure...

SL 2 : *performances meilleures* → résultat d'une épreuve, d'un test → Donc de qualité supérieure...

Bonne performance

SL 1 – noyau 2 : *bonne performance* → rendement maximal, fiabilité élevée → Donc positif, utile, agréable, efficace...

SL 2 : *bonne performance* → résultat d'une épreuve, d'un test → Donc positif, utile, agréable, efficace...

L'application du TSA montre que l'adjonction du collocatif à la base *performance(s)* forme un déploiement argumentatif possible (H32.1). Par conséquent, nous concluons à un mécanisme de motivation argumentative (H32.2). Concernant *bonne performance*, l'application du TSA montre que l'adjonction du collocatif à la base forme un bloc d'argumentation interne (H31.1) ; cette collocation est donc argumentative (H31.2).

Pour statuer sur leur axiologisation intrinsèque, là encore l'analyse diffère selon la SL mobilisée.

Nous appliquons le TEA à la SL 1 – noyau 2 :

Performance intéressante → Donc progrès

Performance intéressante → *Pourtant progrès

Performance intéressante → *Donc échec

Performance intéressante → Pourtant échec

Les enchaînements normatif « progrès » et transgressif « échec » sont sémantiquement acceptables, ce qui signifie que le PA est intrinsèquement axiologisé positivement.

Nous appliquons le TEA à la SL 2 :

Performance intéressante → Donc progrès

Performance intéressante → *Pourtant progrès

Performance intéressante → *Donc échec

Performance intéressante → Pourtant échec

Dans ce cas, ce sont aussi les enchaînements normatif « progrès » et transgressif « échec » qui sont sémantiquement acceptables, ce qui signifie que l'argumentation est également intrinsèquement axiologisée positivement. De fait, *intéressante*, *supérieure*, *meilleures* et *bonne* étant monovalents positifs, leur adjonction à *performance(s)* confère aux collocations une valeur positive (H32.3). Les collocatifs « contaminent positivement » la bivalence de *performance*. Ce paradigme collocationnel corrobore l'hypothèse formulée pour *baisse de performance*, concernant la contamination de la bivalence de la base le collocatif est axiologiquement marqué.

Au niveau des mécanismes macro-sémantiques d'axiologisation de la signification (H41), rappelons que l'unité lexicale *performance(s)* dans cette acception est un monovalent positif, de même que les collocatifs *intéressante*, *supérieure*, *meilleures* et *bonne*. La juxtaposition au

sein d'une même collocation de deux polarités positives nous permet de statuer sur un mécanisme de renforcement des valeurs axiologiques positives portées par *performance(s)*.

Lorsqu'il s'agit de la SL 2, *performance(s)*, eu égard au phénomène de contamination décrit ci-dessus, au niveau macro-sémantique nous statuons sur un mécanisme d'activation positive du potentiel axiologique bivalent de la base *performance(s)*.

Au niveau discursif, l'étude co- et contextuelle des collocations fait apparaître, d'une part, dans trois quart des cas un mécanisme d'activation de l'axiologisation positive, le plus souvent porté par les unités *atteindre*, *obtenir* ou encore *très*, d'autre part, un mécanisme de neutralisation du potentiel axiologique positif, avec des mots comme *présenter*, *avoir* ou encore *montrer*.

Activation

Ex : *les systèmes de reconnaissance automatique de la parole (RAP) actuels atteignent des **performances intéressantes** dans des applications ciblées* (AC_0080)

Ex : *l'ensemble de ces méthodes atteint des **performances supérieures** à eYfhg* (AC_0080)

Ex : *les classifieurs bayésiens de type unigrammes obtiennent les **meilleures performances** d'identification thématique* (AC_0050)

Ex : *ces systèmes ont de très **bonnes performances** et commencent à être déployés en environnement industriel* (AC_0101)

Neutralisation

Ex : *ces approches présentent des **performances intéressantes** sur des textes* (AC_0078).

Ex : *il était possible de produire une combinaison de systèmes ayant des **performances meilleures** sur l'une des dimensions (précision ou décision)* (AC_0117)

Ex : *les résultats de notre test sur un corpus de 500 énoncés ont montré la **bonne performance** à détecter et à corriger les extragrammaticalités, ainsi que l'efficacité de notre approche à éviter les cas de surgénérativité* (AC_0121)

1.2.3.2.2 Les bivalents

Le TSA concerne les deux collocations *performance globale* et *performance réelle*.

Performance globale

SL 1 – noyau 2 : *performance globale* → rendement maximal, fiabilité élevée → Donc considéré dans son ensemble... Donc complet, intégral...

SL 2 : *performance globale* → résultat d'une épreuve, d'un test → Donc considéré dans son ensemble... Donc complet, intégral...

Performance réelle

L'adjectif *réel* étant placé après le nom, il perd la possibilité de sa monovalence positive. Placé avant le nom, *réelle performance* aurait signifié *performance évidente*, *notable*. Placé après le nom, *performance réelle* se rapproche de *performance effective*, c'est pourquoi il est analysé en qualité de bivalent.

SL 1 – noyau 2 : *performance réelle* → rendement maximal, fiabilité élevée → Donc effectif, mis en œuvre...

SL 2 : *performance réelle* → résultat d'une épreuve, d'un test → Donc effectif, mis en œuvre...

Pour les deux SL, l'application du TSA montre que l'adjonction du collocatif à la base *performance(s)* forme un déploiement argumentatif possible (H32.1). Par conséquent, nous concluons à un mécanisme de motivation argumentative (H32.2).

L'axiologisation intrinsèque (H32.3) opère différemment selon la SL de *performance(s)* qui est mobilisée.

Nous appliquons le TEA à la SL – noyau 2 :

Performance globale → Donc progrès

Performance globale → *Pourquoi progrès

Performance globale → *Donc échec

Performance globale → Pourquoi échec

Seuls les enchaînements normatif « progrès » et transgressif « échec » sont sémantiquement acceptables. De fait, *performance* est monovalent positif, ce qui contraint l'instanciation positive de l'unité bivalente *globale*. La polarité positive de la base contamine la bivalence du collocatif. La collocation est donc intrinsèquement axiologisée positivement.

Nous appliquons le TEA à la SL 2 :

Performance globale → Donc progrès

Performance globale → Pourquoi progrès

Performance globale → Donc échec

Performance globale → Pourquoi échec

Dans ce cas, la base et le collocatif sont tous deux bivalents, c'est pourquoi tous les enchaînements, qu'ils soient normatifs ou transgressifs, sont sémantiquement acceptables. La collocation conserve donc son potentiel axiologique intrinsèque, sans que l'un ou l'autre des pôles positif ou négatif ne soient activés.

Au niveau macro-sémantique, dans le cas de la SL 1 – noyau 2, l'axiologisation positive du bivalent résultant de la monovalence positive de la base, nous statuons sur un mécanisme d'activation de l'axiologisation de la signification.

Dans le cas de la SL 2, nous statuons sur un mécanisme de neutralisation des potentiels axiologiques, dans la mesure où ni les éléments de signification de la base ni ceux des collocatifs ne déclenchent d'orientation axiologique. Seul un mécanisme discursif peut rompre la bivalence « globale » de la collocation !

Au niveau discursif, les informations cotextuelles montrent deux types de mécanismes discursifs, sans que l'on puisse les regrouper autour de l'une ou l'autre des significations mobilisées. Un premier mécanisme d'activation positive de l'axiologisation du sens, engendré par *améliorer*, rencontré à plusieurs reprises, et un second mécanisme discursif de neutralisation, porté par des unités elles-mêmes bivalentes comme *juger* ou bien par des mots dont la polarité est marquée, mais dont le cumul annihile les pouvoirs contaminant, dans notre exemple *difficile d'apprécier*.

Activation

Ex : cette liste (probabilisée) de séquences était alors insérée à un décodeur afin d'améliorer les **performances globales** de traduction (AC_0063)

Neutralisation

Ex : *pour ces types d'applications, le taux d'entités nommées correctement identifiées est bien souvent un facteur primordial pour juger des **performances globales** de l'application* (AR_0034)

Ex : *il est difficile d'apprécier avec précision les **performances réelles** de ces systèmes d'identification de la langue* (AR_0034)

1.2.3.2.3 Le neutre

Le TSA est appliqué à *performance moyenne*. Nous avons choisi de catégoriser *moyenne* comme unité neutre sur la base des éléments de SL « ni bon, ni mauvais » ; la SL aurait mentionné « soit bon, soit mauvais », nous l'aurions catégorisée en bivalent.

Performance moyenne

Avant de commencer l'analyse de cette collocation, nous avons pris soin de vérifier qu'il ne s'agissait d'un terme complexe, au même titre que *écart type*, par exemple, mais nous n'en avons trouvé aucune mention lexicographique.

SL 1 – noyau 2 : *performance moyenne* → rendement maximal, fiabilité élevée → *Donc ni bon, ni mauvais...

SL 1 – noyau 2 : *performance moyenne* → rendement maximal, fiabilité élevée → Pourtant ni bon, ni mauvais...

SL 2 : *performance moyenne* → résultat d'une épreuve, d'un test → *Donc (évaluation) ni bon, ni mauvais...

SL 2 : *performance moyenne* → résultat d'une épreuve, d'un test → Pourtant (évaluation) ni bon, ni mauvais...

Pour les deux SL, l'application du TSA montre que l'adjonction du collocatif *moyen* à la base *performance* forme un déploiement argumentatif transgressif possible (H32.1). Nous concluons à un mécanisme de motivation argumentative (H32.2).

L'axiologisation intrinsèque de *performance moyenne* varie selon la SL mobilisée.

Nous appliquons le TEA à la SL – noyau 2 (H32.3) :

Performance moyenne → Donc progrès

Performance moyenne → *Pourtant progrès

Performance moyenne → *Donc échec

Performance moyenne → Pourtant échec

Seuls les enchaînements normatif « progrès » et transgressif « échec » sont sémantiquement acceptables, ce qui signifie que la collocation est intrinsèquement axiologisée positivement. L'orientation axiologique positive de la base *performance* contraint l'instanciation positive de l'ensemble de la collocation, sans que nous ayons la preuve qu'elle « contamine » la neutralité des collocatifs.

Nous appliquons le TEA à la SL 2 (H32.3) :

Performance moyenne → Donc progrès

Performance moyenne → Pourtant progrès

Performance moyenne → Donc échec

Performance moyenne → Pourtant échec

Dans ce cas, la base est bivalente et le collocatif est neutre. Là encore, tous les enchaînements, qu'ils soient normatifs ou transgressifs, sont sémantiquement acceptables. La collocation est donc porteuse d'un potentiel axiologique intrinsèque, sans que l'un ou l'autre des pôles positif ou négatif ne soient activés.

Au niveau macro-sémantique, dans le cas de la SL 1 – noyau 2, l'axiologisation positive de la collocation résultant de la monovalence positive de la base, nous statuons sur un mécanisme d'activation de l'axiologisation de la signification.

Dans le cas de la SL 2, nous statuons sur un mécanisme de neutralisation des potentiels axiologiques. Seul un mécanisme discursif peut rompre la bivalence de la collocation.

De fait, au niveau discursif, l'observation co- et contextuelle de la collocation montre une grande complexité des mécanismes axiologiques en présence. Dans les extraits trouvés, le cotexte semble activer dans la majorité des cas un potentiel axiologique positif, avec des formes linguistiques toutefois complexes. Dans le premier exemple, la juxtaposition de *permettre*, *accroître* et *légèrement* nous pousse à conclure à un mécanisme d'activation de l'axiologisation positive de *performance moyenne*, de même que pour le deuxième extrait. Sans *légèrement*, nous aurions statué sur un mécanisme de renforcement. Le troisième exemple introduit la collocation avec un verbe on ne peut plus neutre, *avoir*. Cependant, les informations contenues entre parenthèses sont extrêmement chargées axiologiquement. Nous considérons dans ce cas que *bien meilleures* l'emporte sur *mais insuffisantes*, c'est pourquoi nous statuons également sur un mécanisme d'activation d'un potentiel axiologique positif. Enfin, dans le dernier extrait, le cotexte est aussi axiologiquement riche ; l'adjonction de la négation *ne – que* au verbe *détériorer* suivie de l'adverbe *peu* complique sensiblement l'interprétation de l'énoncé, que nous ramenons toutefois à un mécanisme d'affaiblissement.

Activation

Ex : *le recours à une sélection optimale (le système connaissant toujours toutes les serveurs possédant au moins un document pertinent) permet d'accroître légèrement la **performance moyenne**.* (AC_0119)

Ex : *ce type de connaissances permet de développer un processus d'analyse réalisant une segmentation des textes avec des **performances moyennes*** (AR_0005)

Ex : *les deux évaluations ont des **performances moyennes** (bien meilleures qu'un classement aléatoire mais insuffisantes pour une utilisation réelle)* (AC_0101)

Affaiblissement

Ex : *le score ne détériore que peu la **performance moyenne** (perte moyenne de 4,61 %)* (AC_0119)

1.3 Synthèse sur les monovalents positifs

Cette synthèse repose sur les analyses sémantico-discursives menées, concernant le binôme monovalent positif *améliorer-amélioration(s) – performance(s)*. Ces deux bases rassemblent à elles deux 33 collocations différentes et 417 occurrences.

Les monovalents positifs ont généré quelques collocations argumentatives, parmi lesquelles *gain de performance*, *baisse de performance*, *dégradation des performances*, etc. Pour toutes ces collocations, l'application du TSA a montré que l'adjonction du collocatif à la base forme avec les stéréotypes proposés ou envisageables de *améliorer-amélioration(s)* ou

performance(s) un bloc d'argumentation interne (H31.1). Aussi, toutes ces collocations correspondent à la concrétisation d'un PA dans le discours (H31.2) ; elles sont prévisibles, « calculables » sur la base des PA existant pour chaque mot. Nous rappelons que de notre point de vue, ces collocations ne forment pas un bloc d'argumentation externe au sens où aucun connecteur argumentatif n'articule le collocatif à la base, elles correspondent à la réalisation de surface d'un bloc d'argumentation externe sous-jacent.

Conformément aux postulats théoriques de la SPA-ALD, de par le lien argumentatif posé, les PA sont par essence évaluatifs. Le TEA a montré que toutes les collocations argumentatives relevées sont axiologisées soit positivement comme pour *bonne performance*, soit négativement comme pour *diminution des performances*.

Le PA est axiologisé positivement lorsque le collocatif, qu'il s'agisse d'un nom ou d'un adjectif antéposé par rapport à la base, est également positif. Dans ce cas, l'adjonction de deux polarités positives crée un mécanisme macro-sémantique de renforcement (H41), comme pour *gain de performance* par exemple.

Le PA est axiologisé négativement lorsque le collocatif nominal, antéposé par rapport à la base, est négatif. Dans ce cas, l'adjonction de deux polarités opposées crée un mécanisme macro-sémantique d'interversion (H41), comme pour *dégradation de performance*.

Deux mécanismes discursifs ont été rencontrés : la neutralisation (**Ex** : *évaluer* le gain en performance) et l'affaiblissement (**Ex** : *faible* gain en performance).

Les monovalents positifs ont également généré plusieurs collocations argumentativement motivées, parmi lesquelles *amélioration de la cohérence*, *amélioration des performances*, *amélioration sensible*, etc. Pour toutes les collocations concernées, l'application du TSA a montré que l'adjonction du collocatif à la base *améliorer-amélioration(s)* ou *performance(s)* forme un déploiement argumentatif possible (H32.1). Aussi, toutes ces collocations correspondent à un mécanisme de motivation argumentative (H32.2). Elles entretiennent une relation argumentative avec la base, ce qui permet de décrire et d'expliquer le principe sous-jacent à l'association des deux entités formant la collocation. Le mécanisme de motivation argumentative revêt un caractère moins prédictible que celui mobilisé par les collocations argumentatives. Toutefois, l'émergence de principes associatifs récurrents en fonction de la monovalence (positive ou négative) ou de la bivalence de la base est possible. De même que pour les collocations argumentatives, nous trouvons plus juste de parler de réalisation de surface d'un bloc d'argumentation externe sous-jacent.

Conformément aux postulats théoriques de la SPA-ALD, de par le lien argumentatif déployé, ces collocations comportent nécessairement une évaluation, orientée vers l'un ou l'autre des pôles axiologiques, positif ou négatif (H32.3). Le TEA a montré que pour tous les cas de motivation argumentative, l'orientation axiologique des collocations est positive comme pour *amélioration de la pertinence*, *amélioration significative*, etc.

La collocation est axiologisée positivement lorsque le collocatif, qu'il s'agisse d'un nom ou d'un adjectif postposé par rapport à la base, est positif ou bivalent. De même que la collocation est également positive, lorsque le collocatif adjectif antéposé est négatif. En revanche, le mécanisme macro-sémantique varie en fonction de la nature axiologique du collocatif :

- quand le collocatif est positif, l'adjonction de deux polarités positives crée un mécanisme de renforcement (H41), comme pour *amélioration substantielle* ;

- quand le collocatif est bivalent, l'adjonction de la polarité positive de la base semble contraindre l'orientation positive du collocatif créant un mécanisme d'activation (H41), comme pour *amélioration de la qualité* ;
- quand le collocatif est négatif, l'antéposition de la polarité négative sur la polarité positive de la base crée un mécanisme d'affaiblissement, comme pour *légère amélioration*.

Quatre mécanismes discursifs ont été rencontrés : le renforcement (**Ex** : *produire une nouvelle* amélioration de la pertinence), l'activation (**Ex** : *source* d'amélioration de la pertinence), la neutralisation (**Ex** : *montrer* une amélioration des performances) et l'interversion (**Ex** : *non*-amélioration des résultats).

Enfin, nous avons trouvé deux collocations arbitraires formées sur la base *améliorer-amélioration* : *amélioration de la méthode* et *amélioration des systèmes*. La collocation arbitraire correspond au cas où la collocation repérée ne correspond ni à la concrétisation d'un PA dans le discours, ni à un cas de motivation argumentative. Dans la mesure où l'adjonction du collocatif *méthode* à la base *amélioration* ne forme pas avec ses stéréotypes proposés et envisageables un bloc d'argumentation interne ou un déploiement argumentatif possible (H33.1), la collocation *amélioration de la méthode* est qualifiée d'arbitraire (H33.2), au sens où elle ne peut être sémantiquement prédictible ; elle n'a de réalité que discursive.

Par l'absence de lien argumentatif, la collocation arbitraire ne devrait pas être intrinsèquement axiologisée, ce que nous discutons. Le TEA a montré que les deux collocations étaient axiologisées positivement (H33.3), sans être intrinsèquement argumentatives, tout en conservant leur potentiel axiologique. La collocation est axiologisée positivement lorsque le collocatif nominal, postposé par rapport à la base, est également neutre. Dans ce cas, la polarité positive de la base semble « contaminer » la neutralité du collocatif, créant un mécanisme d'activation (H41).

Deux mécanismes discursifs ont été rencontrés : le renforcement (**Ex** : *les coûts sont réduits* par l'amélioration de la méthode) et l'activation (**Ex** : *réussir* à améliorer le système).

Pour statuer d'un point de vue un peu plus général sur l'hypothèse H42 relative à la prépondérance des mécanismes discursifs de renforcement et d'activation dans le discours scientifique spécialisé, nous confrontons les mécanismes observés sur les trois ensembles de collocations. Il en ressort que :

- quand la collocation est axiologisée positivement (mécanisme macro-sémantique de renforcement ou d'activation), les mécanismes discursifs correspondent plutôt au renforcement et à l'activation ;
- quand la collocation est axiologisée négativement (mécanisme macro-sémantique d'interversion), les mécanismes discursifs correspondent au renforcement ou à la neutralisation.

2 Les monovalents négatifs

2.1 Erreur(s)

2.1.1 Dépouillement des données

2.1.1.1 Fréquences et nettoyage

Le corpus TAL(N) comporte 148 occurrences d'*erreur* au singulier et 380 occurrences du mot au pluriel.

Selon le critère statistique, nous éliminons les candidats trop peu fréquents. Nous retirons donc tous les apax, les candidats de fréquence deux et les cooccurents n'apparaissant pas dans trois articles distincts.

Ex : *erreur d'appréciation* (1), *erreur non négligeable* (2), *erreur(s) résiduelle(s)* (3, AC_0117), *commettre l'erreur* (1), *erreurs conceptuelles* (1).

Selon le critère de pertinence, nous laissons de côté toutes les collocations grammaticales et autres constructions dont le collocatif ne présente pas de contenu sémantique.

Ex : *erreur(s) due(s) à/aux autres erreurs* (7), *certaines erreurs* (8).

Selon le critère lexical, aucune lexie construite sur *erreur(s)* n'apparaît dans le corpus TAL(N).

Selon le critère lexicographique, nous avons relevé trois termes complexes construits sur *erreur(s)* : *correction d'erreur(s)* (30), *détection d'erreur(s)* (26) et *taux d'erreur* (51).

Concernant *correction d'erreur(s)*, le GDT propose deux définitions spécialisées au domaine de l'informatique dont l'une, synonyme de *nettoyage de données*, correspond à l'emploi qui en est fait dans le corpus. Le GDT notifie qu'avec l'avènement des techniques d'exploration de données et d'extraction de connaissances cette opération a une importance de plus en plus grande et va bien au-delà de la simple correction ou suppression d'erreurs d'ordre technique. La correction d'erreur(s) porte sur la validité des données premières, sur lesquelles reposent la qualité et la fiabilité de l'information et des connaissances qui en sont tirées, ce qui explique que cette activité semble actuellement à l'origine d'un *domaine* (AC_0044) qui développe ses propres *algorithmes* et *modèles* (AR_0049, AC_0011) et cherche à en évaluer les *coûts* (AC_0117).

Nous avons également trouvé dans le GDT une entrée restreinte au domaine informatique pour *détection d'erreurs* et *taux d'erreur(s)*.

Ex : *le TAL en général et la détection d'erreurs en particulier subissent l'influence du contexte dans lequel les outils vont être utilisés* (AR_0008)

2.1.1.2 Liste des collocations lexicales

Cette sous-section présente les quatre tableaux récapitulatifs des collocations lexicales construites sur erreur(s). Les deux premiers tableaux correspondent au contexte droit et les deux derniers, au contexte gauche du substantif.

COLLOCATIONS LEXICALES	FREQ.	AUTRES FORMES	FREQ.	TOTAL
Erreur d'accord	3	-Erreurs d'accord(s)	10	13
Erreur d'étiquetage	5	-E. de l'étiquetage -Erreurs d'étiquetage	1 9	15
Erreur fréquente	-	-E. [...] fréquente -Erreurs fréquentes -Erreurs [...] fréquentes -Type d'E. fréquent	1 1 3 2	7
Erreur de reconnaissance	2	-Erreurs de reconnaissance -Erreurs de la reconnaissance	9 2	13
Erreur du système	1	-Erreurs du système -Erreurs des systèmes	1 1	3

Tableau 12 : Collocations lexicales contexte droit de erreur(s)

COLLOCATIONS LEXICALES	FREQ.	AUTRES FORMES	FREQ.	TOTAL
Erreurs d'analyse	6	-E. [...] de l'analyse	2	8
Erreurs de segmentation		-	-	6

Tableau 13 : Collocations lexicales contexte droit de erreurs

COLLOCATIONS LEXICALES	FREQ.	AUTRES FORMES	FREQ.	TOTAL
Diagnostic d'erreur	1	-Le diagnostic d'erreurs	3	4
Pourcentage d'erreur	2	-Le pourcentage d'erreurs	3	5
Réduction d'erreur	2	-Réduire les erreurs	3	5
Repérage d'erreur	3	-Le repérage d'erreurs	2	5
Risque d'erreur	5	-Le risque d'erreurs	1	6

Tableau 14 : Collocations lexicales contexte gauche de erreur(s)

COLLOCATIONS LEXICALES	FREQ.	AUTRES FORMES	FREQ.	TOTAL
Analyse des erreurs	4	-Analyse [...] des erreurs	4	8
Diminution d'erreurs	1	-Diminuer les erreurs	3	4
Eventuelles erreurs	5	-	-	5
Evitement des erreurs	1	-Eviter les erreurs	5	6

Tableau 15 : Collocations lexicales contexte gauche de erreurs

2.1.2 Signification lexicale de erreur

Dans le LL, *erreur* bénéficie d'une entrée à part entière, dont l'article se décompose en un ensemble de significations du nom au singulier et une signification attachée au nom pluriel (cf. Annexe 6).

Au singulier, on compte six définitions dont l'une consacrée à une lexie restreinte au domaine de la physique – *erreur absolue* –, quatre lexies – *erreur judiciaire* ; *erreur n'est pas compte* ; *erreur relative* ; *erreur systématique* – et une locution adverbiale – *par erreur*. Parmi les cinq définitions restantes, les trois premières sont générales et correspondent à trois noyaux différents, nous les conservons donc pour la construction de la SL de *erreur(s)*. Les deux dernières sont spécialisées, l'une au domaine de la métrologie, science des poids et mesures, l'autre au domaine du droit ; nous les laissons donc de côté.

Définition 1 : Action de se tromper, faute commise en se trompant (*l'erreur est la règle* → bêtise, méprise ; *erreur de calcul* → inexactitude).

Définition 2 : État de qqn qui se trompe (*être dans l'erreur, persister dans son erreur, induire en erreur*).

Définition 3 : Action faite mal à propos, inconsidérée, regrettable (*cette intervention fut une erreur* → gaffe, maladresse ; *erreur de jeunesse* → égarement).

Au pluriel, l'acception du mot est générale, entendue comme « actes condamnables, dérèglements » (*avoir des erreurs à se faire pardonner*).

Le GR propose une entrée pour *erreur* (cf. Annexe 6). L'article, assez dense (3 pages), se décompose en trois catégories de significations : générale, objective et restreinte au sens d'errer ça et là. Naturellement, nous laissons cette dernière catégorie de côté. Dans la catégorie générale, on compte six définitions et trois locutions – *faire erreur* ; *par erreur* ; *sauf erreur*. Concernant les définitions, ce qui caractérise le GR, c'est leur multiplication presque excessive. Par exemple, la première définition proposée est identique à la deuxième, à la différence que dans la deuxième définition est mentionnée une possibilité de construction grammaticale avec *dans*, *en* et *de*. Aussi, nous avons choisi de les regrouper.

Définition 1 : Acte de l'esprit qui tient pour vrai ce qui est faux et inversement.

→ errement, faute, fourvoiement (*erreur choquante/grossière*)

→ bêtise, confusion, malentendu, méprise (*erreur capitale/totale* ; *profonde erreur*)

→ abuser, s'égarer (*faire/commettre une erreur*)

Nous avons laissé de côté deux définitions générales, l'une au sens de *mensonge*, l'autre au sens de *contre-vérité*, et une définition absolue où *erreur*, employée uniquement au singulier, correspond à : *hérésie*, pour leur mention restreinte au GR.

Définition 2 : Action regrettable, maladroite, déraisonnable.

→ étourderie, gaffe, impair, maladresse (*erreur de savoir-vivre* ; *erreur de tactique*)

→ défaillance, faiblesse (*erreurs de jeunesse* ; *erreurs passées*)

Dans la catégorie objective, on compte deux définitions et une lexie – *erreur judiciaire*. De même que précédemment, les deux définitions sont extrêmement proches ; aussi, nous avons décidé de les regrouper.

Définition 3 : Chose fausse, erronée, élément inexact par rapport à une norme, une opération particulière.

→ faute, inexactitude (*relever une erreur*)

→ corriger, effacer, rectifier (*erreur typographique* ; *erreur de calcul*)

Les deux dictionnaires décrivent *erreur* comme un nom féminin. Cette unité lexicale appartient à la zone modale de l'axiologique négatif, conformément aux synonymes proposés, ce qui fait de lui un monovalent négatif.

Figure ci-dessous le tableau de synthèse de la signification lexicale du substantif.

ERREUR(S)			
TRAITS NECESSAIRES DE CATEGORISATION (TNC)			
TYPE GRAMMATICAL		ZONE MODALE	ORIENTATION AXIOLOGIQUE
Nom – Féminin		Axiologique	Monovalent – Négatif
ENTITES SEMANTIQUES			ENTITES DISCURSIVES
NOYAU		STEREOTYPES	POSSIBLES ARGUMENTATIFS
1	Résultat : chose fausse, élément inexact par rapport à une norme, à une opération	Donc réalisation, évaluation... Donc correction, suppression... Donc rectification, révision...	E. donc faute commise E. donc correction typographique E. donc rectification du calcul
2	Etat de qqn qui tient pour vrai ce qui est faux et inversement	Donc fourvoiement, méprise... Donc confusion, malentendu... Donc égarement...	E. donc lourde méprise E. donc grave confusion E. donc profond égarement
3	Action regrettable, inconsidérée, maladroite, déraisonnable	Donc gaffe, maladresse... Donc faiblesse, défaillance...	E. donc grossière maladresse E. faiblesse de jeunesse

Tableau 16 : Signification lexicale de erreur(s)

2.1.3 Analyse des données

2.1.3.1 Le collocatif est un nom

2.1.3.1.1 Les monovalents positifs

Le TSA concerne les collocations *repérage d'erreur* et *analyse des erreurs*.

Repérage d'erreur

SL – noyau 1 : *repérage d'erreur* → chose fausse, élément inexact → *Donc détection, reconnaissance, identification, localisation...

SL – noyau 1 : *repérage d'erreur* → chose fausse, élément inexact → *Pourtant détection, reconnaissance, identification, localisation...

Analyse des erreurs

Contrairement à *erreur d'analyse* (et *performance d'analyse*), placé avant le nom, *analyse* ne désigne pas par métonymie l'analyseur syntaxique, mais le processus général qui vise à examiner quelque chose pour en rendre compte.

SL – noyau 1 : *analyse des erreurs* → chose fausse, élément inexact → Donc étude détaillée...

Dans le cas de *repérage d'erreur*, l'application du TSA montre que l'adjonction du collocatif à la base ne forme ni bloc d'argumentation interne, ni déploiement argumentatif possible (H33.1). Par conséquent, nous qualifions la collocation d'arbitraire (H33.2). En revanche, dans le cas de *analyse des erreurs*, l'adjonction du collocatif à la base forme avec les stéréotypes proposés ou envisageables de *erreur(s)* un déploiement argumentatif possible (H32.1). Par conséquent, nous statuons sur un mécanisme de motivation argumentative (H32.2).

Pour statuer sur l'axiologisation intrinsèque, nous appliquons le TEA :

Repérage d'erreur → *Donc défaillance du logiciel

Repérage d'erreur → Pourtant défaillance du logiciel

Repérage d'erreur → Donc efficacité du logiciel

Repérage d'erreur → *Pourtant efficacité du logiciel

Analyse des erreurs → *Donc défaillance du logiciel

Analyse des erreurs → Pourtant défaillance du logiciel

Analyse des erreurs → Donc efficacité du logiciel

Analyse des erreurs → *Pourtant efficacité du logiciel

Seuls les enchaînements transgressif « défaillance du logiciel » et normatif « efficacité du logiciel » sont sémantiquement acceptables, ce qui signifie que le déploiement argumentatif est intrinsèquement axiologisé positivement, en l'absence de tout élément co- ou contextuel. De fait, rappelons que l'unité lexicale *erreur* est un monovalent négatif, contrairement aux collocatifs *repérage* et *analyse*, qui sont des monovalents positifs. Contrairement à l'hypothèse posée en cours d'analyse à propos de la polarité de la base supposée « plus forte » que celle du collocatif, ces données montrent que la polarité positive portée par le collocatif l'emporte sur la polarité négative de la base. Nous pensons que ce phénomène est rendu possible d'une part par le statut de substantif du collocatif et d'autre part par la position de tête qu'il occupe au sein de la collocation.

Au niveau des mécanismes macro-sémantiques d'axiologisation de la signification (H41), la contamination de la polarité de la base par celle positive du collocatif produit un mécanisme d'intervention des valeurs axiologiques négatives portées par *erreur(s)*. Nous attribuons le même mécanisme à la collocation *analyse des erreurs*.

Au niveau discursif, l'observation co- et contextuelle des collocations fait apparaître trois mécanismes possibles : l'activation (avec *être utilisable* et *permettre*), la neutralisation (avec *montrer*) et l'affaiblissement (avec *difficile à*).

Activation

Ex : *ces dernières informations sont ainsi utilisables pour le repérage d'erreurs telles que l'omission ou la répétition de segments de texte* (AC_0021)

Ex : *l'analyse des erreurs de participe passé nous permettra de montrer l'interaction des trois approches* (AR_0008)

Neutralisation

Ex : *l'analyse des erreurs* a montré la nécessité d'introduire des éléments syntaxiques (AC_0074)

Affaiblissement

Ex : même dans ce cas, comme présenté plus bas, le **repérage de l'erreur** est difficile à effectuer avec précision (AC_0021)

2.1.3.1.2 Les monovalents négatifs

L'application du TSA concerne les collocations *réduction d'erreur*, *risque d'erreur*, *diminution des erreurs* et *évitement des erreurs*.

Réduction d'erreur

SL – noyau 1 : *réduction d'erreur* → chose fausse, élément inexact → Donc action de diminuer, limiter.... Donc diminué, faible, moins nombreux...

Risque d'erreur

SL – noyau 1 : *risque d'erreur* → chose fausse, élément inexact → *Donc danger plus ou moins prévisible...

SL – noyau 1 : *risque d'erreur* → chose fausse, élément inexact → Pourtant danger plus ou moins prévisible...

Diminution des erreurs

SL – noyau 1 : *diminution des erreurs* → chose fausse, élément inexact → Donc baisse, décroissance, réduction...

Évitement des erreurs

SL – noyau 1 : *évitement des erreurs* → chose fausse, élément inexact → Donc écarter, parer, esquiver⁵...

L'application du TSA montre que l'adjonction du collocatif à la base *erreur(s)* forme un déploiement argumentatif possible (H32.1). Par conséquent, nous concluons à un mécanisme de motivation argumentative (H32.2).

Pour statuer sur leur axiologisation intrinsèque (H32.3), nous appliquons le TEA :

Réduction d'erreur → *Donc défaillance du logiciel

Réduction d'erreur → Pourtant défaillance du logiciel

Réduction d'erreur → Donc efficacité du logiciel

Réduction d'erreur → *Pourtant efficacité du logiciel

Seuls les enchaînements transgressif « défaillance du logiciel » et normatif « efficacité du logiciel » sont sémantiquement acceptables, ce qui signifie que l'argumentation est intrinsèquement axiologisée positivement. L'unité lexicale *erreur* est un monovalent négatif.

⁵ Il aurait été préférable d'employer des substantifs pour le déploiement stéréotypique d'*évitement* ; toutefois, aucun synonyme nominal ne figure dans les dictionnaires.

Contrairement à ce que nous avons pu observer pour la juxtaposition de deux polarités positives (Ex : *amélioration de la cohérence*), dans le cas de deux polarités négatives nous ne pouvons conclure systématiquement à une axiologisation intrinsèque négative. D'un côté, *erreur* est indiscutablement axiologique négatif, de l'autre, une *réduction de la cohérence* (monovalent positif), une *réduction de résultat* (bivalent) et une *réduction de la méthode* (neutre) demeurent également axiologiques négatifs. Pour autant, seule l'axiologisation intrinsèque de *risque d'erreur* est négative, l'axiologisation intrinsèque de chacune des autres collocations est positive. D'où nous concluons que l'adjonction de deux polarités négatives conduit :

- à une axiologisation négative, lorsque l'argumentation déployée est transgressive (**Ex** : *erreurs grossières*, AR_0042) ;
- à une axiologisation positive, lorsque l'argumentation déployée est normative (**Ex** : *réduction d'erreur*).

Au niveau des mécanismes macro-sémantiques d'axiologisation de la signification, il se révèle que l'association de deux polarités négatives engendre :

- soit un phénomène intrinsèque de renforcement (**Ex** : *risque d'erreur*) ;
- soit un phénomène d'interversion des valeurs portées par chaque unité qui compose la collocation (**Ex** : *réduction d'erreur*).

L'étude co- et contextuelle des collocations fait apparaître des mécanismes discursifs différents selon la forme des blocs d'argumentations déployés.

Dans leurs formes normatives, on constate la présence d'un phénomène d'activation (avec *permettre*) et la récurrence du phénomène de neutralisation (avec *indiquer* et *montrer*).

Activation

Ex : *cela nous permet de **réduire les erreurs** de classifications d'entités* (AC_0110)

Neutralisation

Ex : *on indique la **diminution d'erreur** que la GCFG offre en comparaison avec la SCFG* (AC_0069)

Ex : *les résultats de l'évaluation ont montré [...] l'efficacité de cette grammaire pour la désambiguïsation et pour l'**évitement des erreurs** dans l'entrée* (AC_0122)

Dans leur forme transgressive, on relève la possibilité d'un phénomène de neutralisation (avec *être présent*), mais surtout une majorité de phénomènes d'affaiblissement de la polarité négative de la collocation, par la présence notoire dans son cotexte de l'adjectif *faible*, lui-même occasionnellement renforcé par l'adverbe *très*.

Neutralisation

Ex : *mais le **risque d'erreur** est présent partout* (AC_0047)

Affaiblissement

Ex : *cette valeur est peu élevée, mais avec un faible **risque d'erreur*** (AR_0011)

Ex : *tout au moins d'indiquer la branche à suivre avec un **risque d'erreur** très faible* (AC_0050)

2.1.3.1.3 Les bivalents

Le TSA est appliqué à *diagnostic d'erreur*.

Diagnostic d'erreur

SL – noyau 1 : *diagnostic d'erreur* → chose fausse, élément inexact → Donc détermination, évaluation, jugement d'après les signes...

L'application du TSA montre que l'adjonction du collocatif à la base forme avec les stéréotypes proposés de *erreur* un bloc d'argumentation interne (H31.1). Par conséquent, nous concluons pour ces deux collocations à la concrétisation d'un PA dans le discours (H31.2). Il s'agit d'une collocation argumentative formant un bloc d'argumentation externe.

Pour statuer sur leur axiologisation intrinsèque (H31.3), nous appliquons le TEA :

Diagnostic d'erreur → Donc défaillance du logiciel

Diagnostic d'erreur → *Pourtant défaillance du logiciel

Diagnostic d'erreur → *Donc efficacité du logiciel

Diagnostic d'erreur → Pourtant efficacité du logiciel

Seuls les enchaînements normatif « défaillance du logiciel » et transgressif « efficacité du logiciel » sont sémantiquement acceptables, ce qui signifie que le PA est intrinsèquement axiologisé négativement. De fait, rappelons que l'unité lexicale *erreur* est un monovalent négatif, tandis que le collocatif *diagnostic* est bivalent. Bien que situé en position tête, dans ce cas, la bivalence du collocatif est « contaminée » par la polarité négative de la base.

Au niveau macro-sémantique, l'axiologisation négative du bivalent résultant de la monovalence négative de la base *erreur*, nous statuons sur un mécanisme d'activation de l'axiologisation de la signification.

Au niveau discursif, l'observation co- et contextuelle des collocations fait apparaître deux mécanismes possibles : de renforcement (avec *risque de se tromper*) et de neutralisation (avec *établir*).

Renforcement

Ex : *au regard des risques de se tromper dans le diagnostic de l'erreur* (AR_0056)

Neutralisation

Ex : *pour établir un premier diagnostic de ces erreurs* (AR_0054)

2.1.3.1.4 Les neutres

L'application du TSA concerne les collocations *erreur d'accord*, *erreur du système* et *pourcentage d'erreur*.

Erreur d'accord

Dans cette collocation, *accord* est systématiquement employé au sens syntaxique du terme et non dans l'acception qui désigne « un arrangement, une entente entre individus ».

SL – noyau 1 : *erreur d'accord* → chose fausse, élément inexact → *Donc cohésion grammaticale... Donc règle formelle...

SL – noyau 1 : *erreur d'accord* → chose fausse, élément inexact → Pourtant cohésion grammaticale... Pourtant règle formelle...

Erreur du système

SL – noyau 1 : *erreur du système* → chose fausse, élément inexact → *Donc traitements matériels et logiciels de données...

SL – noyau 1 : *erreur du système* → chose fausse, élément inexact → *Pourtant traitements matériels et logiciels de données...

Pourcentage d'erreur

SL – noyau 1 : *pourcentage d'erreur* → chose fausse, élément inexact → *Donc proportion...

SL – noyau 1 : *pourcentage d'erreur* → chose fausse, élément inexact → *Pourtant proportion...

L'application du TSA révèle deux cas de figure. Le premier cas correspond à la collocation *erreur d'accord*. L'adjonction du collocatif à la base *erreur(s)* forme un déploiement argumentatif possible (H32.1). Par conséquent, nous concluons à un mécanisme de motivation argumentative (H32.2).

Pour statuer sur son axiologisation intrinsèque (H32.3), nous appliquons le TEA :

Erreur d'accord → Donc défaillance du logiciel

Erreur d'accord → *Pourtant défaillance du logiciel

Erreur d'accord → *Donc efficacité du logiciel

Erreur d'accord → Pourtant efficacité du logiciel

Seuls les enchaînements normatif « défaillance du logiciel » et transgressif « efficacité du logiciel » sont sémantiquement acceptables, ce qui signifie que l'argumentation est intrinsèquement axiologisée négativement. De fait, rappelons que l'unité lexicale *erreur* est un monovalent négatif, tandis que ce collocatif est neutre. L'orientation axiologique négative de la base *erreur* contraint l'instanciation négative de l'ensemble de la collocation, sans que nous ayons la preuve qu'elle « contamine » la neutralité du collocatif.

Au niveau macro-sémantique, l'axiologisation négative de la collocation résultant de la monovalence négative de la base, nous statuons sur un mécanisme d'activation de l'axiologisation de la signification.

L'étude co- et contextuelle de la collocation fait apparaître deux mécanisme discursifs possibles, l'un de renforcement (avec *mal formulées* et *constructions agrammaticales*), l'autre de neutralisation de la polarité négative (avec *répartir*).

Renforcement

Ex : *l'analyse doit être suffisamment robuste pour reconnaître la structure des phrases même si elles sont mal formulées (i.e. contenant par exemple des **erreurs d'accord** ou des constructions agrammaticales)* (AC_0103)

Neutralisation

Ex : *les 102 **erreurs d'accord** peuvent être réparties entre les quatre structures syntaxiques suivantes* (AR_0108)

Le second cas regroupe les deux autres collocations. Pour *erreur du système*, et *pourcentage d'erreur*, l'application du TSA montre que l'adjonction du collocatif à la base ne forme pas avec les stéréotypes proposés ou envisageables de *erreur(s)* un bloc

d'argumentation interne, ni un déploiement argumentatif possible (H33.1). Par conséquent, ces collocations sont arbitraires (H33.2).

Nous rappelons que, dans ce cas, nous pensons que la base conserve son potentiel axiologique (H33.3). Nous appliquons donc le TEA :

Erreur du système → Donc défaillance du logiciel

Erreur du système → *Pourtant défaillance du logiciel

Erreur du système → *Donc efficacité du logiciel

Erreur du système → Pourtant efficacité du logiciel

Seuls les enchaînements normatif « défaillance du logiciel » et transgressif « efficacité du logiciel » sont sémantiquement acceptables, ce qui signifie que la collocation est intrinsèquement axiologisée négativement, sans être intrinsèquement argumentative, mais tout en conservant son potentiel axiologique. L'orientation axiologique négative de la base *erreur* contraint l'instanciation négative de l'ensemble de la collocation, sans que nous ayons la preuve qu'elle « contamine » la neutralité du collocatif.

Au niveau discursif, l'observation co- et contextuelle des collocations montre un seul mécanisme discursif (H42) de neutralisation (avec *distinguer*, *constituer*, *donner*, *afficher*).

Neutralisation

Ex : *on distingue les erreurs liées à des **erreurs du système** de reconnaissance des erreurs inhérentes à la compréhension seule* (AR_0054)

Ex : *les homophones ou les quasi-homophones constituent l'une des causes principales des **erreurs des systèmes*** (AC_0122)

Ex : *le tableau 3 donne le **pourcentage d'erreur** en substitution (étiquette erronée)* (AC_0039)

2.1.3.2 Le collocatif est un adjectif

2.1.3.2.1 Le monovalent positif

Le TSA est appliqué à la collocation *erreur fréquente*.

Erreur fréquente

SL – noyau 1 : *erreur fréquente* → chose fausse, élément inexact → *Donc réitérée, répétée... *Donc continue, nombreuse, perpétuelle... *Donc commune, courante, habituelle, ordinaire...

SL – noyau 1 : *erreur fréquente* → chose fausse, élément inexact → Pourtant réitérée, répétée... Pourtant continue, nombreuse, perpétuelle... Pourtant commune, courante, habituelle, ordinaire...

L'application du TSA montre que l'adjonction du collocatif à la *erreur(s)* un déploiement argumentatif possible (H32.1). Par conséquent, nous concluons à mécanisme de motivation argumentative (H32.2).

Pour statuer sur son axiologisation intrinsèque (H32.3), nous appliquons le TEA :

Erreur fréquente → Donc défaillance du logiciel

Erreur fréquente → *Pourtant défaillance du logiciel

Erreur fréquente → *Donc efficacité du logiciel

Erreur fréquente → Pourtant efficacité du logiciel

Seuls les enchaînements normatif « défaillance du logiciel » et transgressif « efficacité du logiciel » sont sémantiquement acceptables, ce qui signifie que le déploiement argumentatif est intrinsèquement axiologisé négativement. L'unité lexicale *erreur* est un monovalent négatif, contrairement au collocatif *fréquente*, qui est un monovalent positif. Conformément à l'hypothèse posée en cours d'analyse à propos de la polarité de la base supposée « plus forte » que celle du collocatif, ces données montrent que la polarité négative portée par la base l'emporte sur la polarité positive du collocatif. Notons cependant que la base occupe la position tête de la collocation et que le collocatif est un adjectif.

Au niveau du mécanisme macro-sémantique d'axiologisation de la signification (H41), la contamination de la polarité positive du collocatif par celle négative de la base produit un mécanisme de renforcement des valeurs axiologiques négatives portées par *erreur(s)*.

Au niveau discursif, l'étude co- et contextuelle fait apparaître uniquement un mécanisme de neutralisation du potentiel axiologique négatif, porté dans les extraits ci-dessous par les unités *concerner* et *être*.

Ex : les *erreurs les plus fréquentes* concernent les incises et les faux-départs (AC_0074)

Ex : en fait, les erreurs d'accords sont parmi les *erreurs les plus fréquentes* (AC_0122)

2.1.3.2.2 Le monovalent négatif

Le TSA est appliqué à la collocation *éventuelles erreurs*.

Éventuelles erreurs

SL – noyau 1 : *éventuelles erreurs* → chose fausse, élément inexact → *Donc possible...

*Donc aléatoire, hypothétique, imprévisible, incertaine...

SL – noyau 1 : *éventuelles erreurs* → chose fausse, élément inexact → Pourtant possible...

Pourtant aléatoire, hypothétique, imprévisible, incertaine...

L'application du TSA montre que l'adjonction du collocatif à la base *erreur(s)* forme un déploiement argumentatif possible (H32.1). Nous concluons pour cette collocation à la un mécanisme de motivation argumentative (H32.2).

Pour statuer sur son axiologisation intrinsèque (H32.3), nous appliquons le TEA :

Eventuelles erreurs → Donc défaillance du logiciel

Eventuelles erreurs → *Pourtant défaillance du logiciel

Eventuelles erreurs → *Donc efficacité du logiciel

Eventuelles erreurs → Pourtant efficacité du logiciel

Seuls les enchaînements normatif « défaillance du logiciel » et transgressif « efficacité du logiciel » sont sémantiquement acceptables, ce qui signifie que la collocation est intrinsèquement axiologisée négativement. Conformément à ce que nous avons observé pour *risque d'erreur*, l'adjonction de deux polarités négatives conduit à une axiologisation négative, lorsque l'argumentation déployée est transgressive.

Au niveau du mécanisme macro-sémantique d'axiologisation de la signification, nous statuons sur un mécanisme de renforcement de la polarité négative de la base.

Au niveau discursif, l'étude co- et contextuelle fait apparaître uniquement un mécanisme de neutralisation du potentiel axiologique négatif, porté dans les extraits ci-dessous par les unités *localisation* et *prise en charge*.

Neutralisation

Ex : *elle ne donne cependant pas d'indication quant à la localisation des éventuelles erreurs repérées* (AC_0021)

Ex : *la prise en charge des éventuelles erreurs (e.g. erreurs d'accord, erreurs de structure...) est totalement intégrée au processus d'analyse* (AC_0103)

2.2 Problème(s)

2.2.1 Dépouillement des données

2.2.1.1 Fréquences et nettoyage

Le corpus TAL(N) comporte 552 occurrences de *problème* au singulier et 336 occurrences du mot au pluriel.

Selon le critère statistique, nous éliminons les candidats trop peu fréquents. Nous retirons donc tous les apax, les candidats de fréquence deux et les cooccurrents n'apparaissant pas dans trois articles distincts.

Ex : *problème de calcul* (1), *problème trivial* (2), *ampleur du problème* (1), *catégories de problèmes* (3, AC_0018), *problème passionnant* (1).

Selon le critère de pertinence, nous laissons de côté toutes les collocations grammaticales et autres constructions dont le collocatif ne présente pas de contenu sémantique.

Ex : *problème(s) à/de*, *type(s) de problèmes* (6).

Selon le critère lexical, une seule lexie (appelée locution dans le GR) construite sur *problème(s)* est présente dans le corpus TAL(N) ; il s'agit de *poser des problèmes*, défini par – susciter des difficultés qu'il faudra résoudre –, qui « occure » environ une centaine de fois⁶.

Aucun terme complexe, construit sur *problème(s)* et entrant dans la catégorie du critère lexicographique, n'a été relevé dans le corpus.

2.2.1.2 Liste des collocations lexicales

Figurent ci-après les trois tableaux récapitulatifs des collocations construites sur *problème(s)*.

⁶ Les formes de cette locution sont tellement variées (*poser problème*, *pose tellement de problèmes*, *ne posent aucun problème*, *problèmes posés...*) que même avec une recherche semi-automatisée en corpus il est difficile d'avancer un nombre plus précis.

COLLOCATIONS LEXICALES	FREQ.	AUTRES FORMES	FREQ.	TOTAL
Problème classique	4	-P. [...] classique -Classique P. -Problèmes classiques	1 1 1	7
Problème complexe	2	-P. [...] complexe	1	3
Problème crucial	4	-P. [...] crucial -Problèmes cruciaux	2 2	8
Problème d'accord	2	-P. de l'accord -Problèmes d'accord	1 1	4
Problème d'ambiguïté	1	-P. de l'ambiguïté -Problèmes d'ambiguïté(s)	1 3	5
Problème de cohérence	2	-Problèmes de cohérence	2	4
Problème de compréhension	2	-P. de la compréhension	1	3
Problème de désambiguïsation	1	-P. de la désambiguïsation -Problèmes de désambiguïssations	2 1	4
Problème de la polysémie	3	-Problèmes de polysémies	1	4
Problème de la sélection	3	-P. de sélection -Problèmes de la sélection -Problèmes de sélection	1 2 2	8
Problème de l'analyse	3	-Problèmes d'analyse	7	10
Problème de l'évaluation	6	-P. d'évaluation -Problèmes d'évaluation	1 1	8
Problème de segmentation	2	-Problèmes de segmentation	1	3
Problème difficile	3	-P. [...] difficile -Difficiles problèmes	6 1	10
Problème du choix	5	-	-	5
Problème essentiel	4	-Problèmes essentiels	1	5
Problème épineux	1	-Problèmes épineux -Epineux problèmes	2 2	5
Problème général	3	-P. [...] général -Problèmes généraux	2 3	8
Problème important	4	-Important problème -Problèmes importants -Problèmes [...] importants -Importants problèmes	2 3 1 2	12
Problème majeur	8	-Problèmes majeurs	5	13
Problème particulier	6	-Problèmes particuliers	3	9
Problème pratique	1	-Problèmes pratiques	4	5
Problème principal	7	-Principal problème -Problèmes principaux -Principaux problèmes	1 1 4	13
Problème sémantique	1	-Problèmes sémantiques	2	3
Problème spécifique	2	-Problèmes spécifiques	6	8
Problème technique	3	-Problèmes techniques	2	5

Tableau 17 : Collocations lexicales contexte droit de problème(s)

COLLOCATIONS LEXICALES	FREQ.	AUTRES FORMES	FREQ.	TOTAL
Problèmes linguistiques	12	-	-	12

Tableau 18 : Collocations lexicales contexte droit de problèmes

COLLOCATIONS LEXICALES	FREQ.	AUTRES FORMES	FREQ.	TOTAL
Réponse au problème	3	-Réponse aux problèmes -Répondre au problème -Répondre aux problèmes	3 2 4	12
Résolution du problème	3	-Résolution(s) des problèmes -Résoudre [...] problème -Résoudre [...] problèmes	8 27 21	59
Solution au problème	-	-Solution [...] au problème -Solution(s) aux problèmes	7 9	16
Traitement du problème	1	-Traiter le problème -Traiter les problèmes	11 9	21
Vrais problèmes	3	-	-	3

Tableau 19 : Collocations lexicales contexte gauche de problème(s)

2.2.2 Signification lexicale de problème

Dans le LL, *problème* bénéficie d'une entrée à part entière sous laquelle figure également problématique ; problématisation (cf. Annexe 7). L'article propose seulement trois définitions, dont l'une familière, consacrée à la lexis – Il n'y a pas de problème –, que nous laissons de côté. Les deux autres sont générales, quoique l'une d'entre elles est d'orientation plutôt scientifique ; nous les utilisons pour la construction de la SL.

Définition 1 : (Scientifique) Question qui appelle une solution d'ordre logique, rationnel (*problème mathématique ; données du problème*).

Définition 2 : Tout ce qui est difficile à résoudre, à expliquer (*problème moral, problème technique, problème de circulation*).

Le GR propose également une entrée pour *problème* (cf. Annexe 7). L'article se compose de trois définitions générales. Nous conservons les deux premières ; en revanche, la troisième se rapporte à la deuxième par métonymie – Chose, personne qui pose un, des problèmes ; chose incompréhensible (énigme → *son attitude est un problème*) –, pour sa mention restreinte au GR.

Définition 1 : Question à résoudre, point obscur que l'on se propose d'éclaircir, qui prête à discussion dans un domaine quelconque de la connaissance.

→ question, traiter, résoudre (*problème philosophique, soulever le problème, problème insoluble*)

→ équation, opération, analyse, aporisme (*problème mathématique ; solution du problème*)

Définition 2 : Difficulté qu'il faut résoudre pour obtenir un résultat ; situation instable ou dangereuse exigeant une décision.

→ ennui, régler (*problème technique ; problème du stationnement*)

Les deux dictionnaires décrivent *problème* comme un nom masculin. Cette unité lexicale appartient à la zone modale de l'axiologique négatif, conformément aux synonymes proposés, ce qui fait de lui un monovalent négatif.

Figure ci-dessous le tableau de synthèse de la signification lexicale du nom.

PROBLEME(S)			
TRAITS NECESSAIRES DE CATEGORISATION (TNC)			
TYPE GRAMMATICAL		ZONE MODALE	ORIENTATION AXIOLOGIQUE
Nom – Masculin		Axiologique	Monovalent – Négatif
ENTITES SEMANTIQUES			ENTITES DISCURSIVES
NOYAU		STEREOTYPES	POSSIBLES ARGUMENTATIFS
1	Processus : question à résoudre qui appelle une solution d'ordre logique, rationnel	Donc poser, soulever, voir... Donc aborder, résoudre, traiter... Donc analyser, déduire, dégager...	P. donc soulever une difficulté P. donc résolution mathématique P. donc déduction impossible
2	Processus : difficulté qu'il faut expliquer ou résoudre pour obtenir un résultat	Donc considérer, conclure... Donc régler, maîtriser... Donc étudier, examiner...	P. donc considération technique P. donc régulation de la circulation P. donc examen méticuleux

Tableau 20 : Signification lexicale de problème(s)

2.2.3 Analyse des données

2.2.3.1 Le collocatif est un nom

2.2.3.1.1 Les monovalents positifs

L'application du TSA concerne les collocations *problème de cohérence*, *problème de la sélection*, *réponse au problème*, *résolution du problème*, *solution du problème* et *traitement du problème*.

Problème de cohérence

SL – noyau 1 : *problème de cohérence* → question à résoudre (par une solution) → *Donc harmonieux, logique, moins contradictoire...

SL – noyau 1 : *problème de cohérence* → question à résoudre (par une solution) → Pourtant harmonieux, logique, moins contradictoire...

SL – noyau 2 : *problème de cohérence* → difficulté à expliquer/résoudre (en vue d'un résultat) → *Donc harmonieux, logique, moins contradictoire...

SL – noyau 2 : *problème de cohérence* → difficulté à expliquer/résoudre (en vue d'un résultat) → Pourtant harmonieux, logique, moins contradictoire...

Problème de la sélection

SL – noyau 1 : *problème de la sélection* → question à résoudre (par une solution) → *Donc (effectuer) meilleur choix possible...

SL – noyau 1 : *problème de la sélection* → question à résoudre (par une solution) → Pourtant (effectuer) meilleur choix possible...

SL – noyau 2 : *problème de la sélection* → difficulté à expliquer/résoudre (en vue d'un résultat) → *Donc (effectuer) meilleur choix possible...

SL – noyau 2 : *problème de la sélection* → difficulté à expliquer/résoudre (en vue d'un résultat) → Pourtant (effectuer) meilleur choix possible...

Réponse au problème

L'analyse cotextuelle montre que pour cette collocation seule la SL – noyau 1 est mobilisée, ce qui est d'une cohérence sémantique absolue compte tenu des éléments de signification du mot *réponse* trouvés dans le GR : « Solution, explication apportée (à une question) par le raisonnement, par un dogme ou une science. »

SL – noyau 1 : *réponse au problème* → question à résoudre (par une solution) → Donc solution, explication apportée par le raisonnement...

Résolution du problème

La présence de *solution* dans les blocs d'argumentations ci-dessous déployés laisse penser au premier abord que seule la SL – noyau 1 est mobilisée. Toutefois, *résolution* est définie dans le GR comme « Action de résoudre, résultat de cette action » et *résoudre* faisant partie intégrante des deux noyaux nous générons les déploiements argumentatifs pour les deux noyaux. En outre, l'analyse cotextuelle de la collocation montre que les deux emplois sont possibles.

Ex : *cette théorie permettrait alors d'envisager la résolution des problèmes* (AR_0057)

Ex : *quoique nos résultats ne soient pas aussi concluants que pour le domaine technique, nous sommes convaincues que la résolution des problèmes d'analyse* (AC_0067)

SL – noyau 1 : *résolution du problème* → question à résoudre (par une solution) → Donc découverte de la solution...

SL – noyau 2 : *résolution du problème* → difficulté à expliquer/résoudre (en vue d'un résultat) → Donc découverte de la solution...

Solution du problème

Les éléments de signification du mot *solution* trouvés dans le GR font apparaître ce qui dans le cadre de la construction de la SL correspondrait à deux noyaux : SL – noyau 1, « Opération mentale qui, en substituant une pluralité analysable à un ensemble complexe d'éléments entremêlés, parvient à surmonter une difficulté, à résoudre une question, un problème théorique ou pratique » ; SL – noyau 2, « Ensemble de décisions et d'actes qui peuvent résoudre une difficulté ». La présence simultanée de *solution* dans la collocation et dans le noyau du mot *problème* nous pousserait théoriquement, comme pour *réponse au problème*, à ne déployer les blocs d'argumentations que pour la SL – noyau 1. Toutefois, au vu de la dualité de la signification de *solution*, non seulement nous décidons de déployer les blocs d'argumentations pour les deux noyaux mais, par souci de rigueur sémantique, nous choisissons également d'y associer les éléments de signification correspondants.

SL – noyau 1 : *solution du problème* → question à résoudre (par une solution) → Donc opération mentale permettant de surmonter une difficulté, résoudre une question, un problème théorique ou pratique...

SL – noyau 2 : *solution du problème* → difficulté à expliquer/résoudre (en vue d'un résultat) → Donc décisions, actes permettant de résoudre une difficulté...

Traitement du problème

Cette collocation étant majoritairement réalisée avec le verbe *traiter*, nous choisissons d'en faire figurer explicitement les éléments de signification dans les déploiements stéréotypiques, d'autant que cette démarche éclaire grandement le statut argumentatif de la collocation.

Le premier bloc d'argumentation mobilise donc la SL de *traitement*, le second, celle de *traiter*.

SL – noyau 1 : *traitement du problème* → question à résoudre (par une solution) → Donc modification... Donc soumettre à l'esprit/la pensée pour étudier/exposer...

SL – noyau 2 : *traitement du problème* → difficulté à expliquer/résoudre (en vue d'un résultat) → Donc modification... Donc soumettre à l'esprit/la pensée pour étudier/exposer...

Pour *cohérence* et *sélection*, l'application du TSA montre que l'adjonction du collocatif à la base forme un déploiement argumentatif possible (H32.1). Par conséquent, nous concluons pour ces deux collocations à un mécanisme de motivation argumentative (H32.2). En revanche, pour les autres collocatifs, l'application du TSA montre que l'adjonction du collocatif à la base forme avec les stéréotypes proposés ou envisageables de *problème(s)* un bloc d'argumentation interne (H31.1). Par conséquent, nous concluons donc à la concrétisation d'un PA dans le discours (H31.2). Il s'agit de collocations argumentatives formant un bloc d'argumentation externe.

Pour statuer sur leur axiologisation intrinsèque, rappelons que l'unité lexicale *problème(s)* est un monovalent négatif, contrairement aux collocatifs mobilisés, qui sont monovalents positifs. Les déploiements argumentatifs permettent de dégager deux cas de figure en fonction de leur caractère normatif ou transgressif.

Nous appliquons le TEA quand *problème* occupe la position tête :

Problème de cohérence → Donc défaillance du logiciel

Problème de cohérence → *Pourquoi défaillance du logiciel

Problème de cohérence → *Donc efficacité du logiciel

Problème de cohérence → Pourquoi efficacité du logiciel

Seuls les enchaînements normatif « défaillance du logiciel » et transgressif « efficacité du logiciel » sont sémantiquement acceptables, ce qui signifie que l'argumentation est intrinsèquement axiologisée négativement. Donc, quand *problème* occupe la position de tête, le bloc d'argumentation déployé est de nature transgressive (*problème de cohérence*, *problème de la sélection*), et l'ensemble de la collocation est axiologisé négativement. La base conserve son potentiel axiologique et contamine le collocatif.

Nous appliquons le TEA quand le collocatif occupe la position tête :

Réponse au problème → *Donc défaillance du logiciel

Réponse au problème → Pourquoi défaillance du logiciel

Réponse au problème → Donc efficacité du logiciel

Réponse au problème → *Pourquoi efficacité du logiciel

Seuls les enchaînements transgressif « défaillance du logiciel » et normatif « efficacité du logiciel » sont sémantiquement acceptables, ce qui signifie que le PA est intrinsèquement axiologisé positivement. Donc, quand le collocatif est en position tête, le bloc d'argumentation déployé est de nature normative (*réponse au problème*, *résolution du problème*, *solution au problème*, *traitement du problème*, *contournement de problème*) et

l'ensemble de la collocation est axiologisé positivement. La base perd son potentiel négatif, désormais positif par la contamination axiologique du collocatif.

Contrairement à l'hypothèse posée en cours d'analyse à propos de la polarité de la base supposée « plus forte » que celle du collocatif, ces données montrent que la polarité positive portée par le collocatif l'emporte sur la polarité négative de la base, quand celui-ci apparaît en position tête et que le bloc d'argumentation déployé est de nature normative.

Au niveau macro-sémantique, dans le premier cas, il s'agit d'un mécanisme de renforcement, tandis que dans le second, il s'agit d'un mécanisme d'interversion.

Au niveau discursif, l'étude des collocations fait apparaître l'existence de quatre mécanismes, dans l'ordre de renforcement (avec *se heurter systématiquement*), de neutralisation (avec *montrer, rencontrer, considérer, consister, proposer* et la juxtaposition de *réponse* et *laisser en suspens*), d'affaiblissement (avec *tentent de résoudre, portée relativement limitée*), d'interversion (avec *proposer une solution, excellents résultats* et *permettre*).

Renforcement

Ex : *une première définition intuitive est toujours possible, mais elle se heurte systématiquement à un **problème de cohérence*** (AR_0007)

Ex : *ces excellents résultats nous permettent de **traiter les problèmes** liés à la capitalisation de textes présentés précédemment* (AR_0042)

Neutralisation

Ex : *l'exemple suivant [...] montre des **problèmes de cohérence*** (AR_0026)

Ex : *on rencontre les **problèmes de sélection** et de fusion* (AC_0119)

Ex : *elle a conduit les chercheurs à considérer la structuration de la terminologie comme une **réponse au problème** de son exploitation* (AR_0066)

Ex : *une **solution aux problèmes** de la méthode précédente consiste à pondérer les critères selon leur pertinence* (AR_0055)

Ex : *cette réponse laisse néanmoins en suspens le **problème de la sélection*** (AR_0074)

Affaiblissement

Ex : *nos travaux en cours tentent de résoudre les mêmes **problèmes de sélection*** (AC_0119)

Ex : *si les deux premières conclusions sont [...] d'une portée relativement limitée dans la **résolution des problèmes** universels* (AC_0062)

Interversion

Ex : *LocalMaxs propose une solution robuste et flexible au **problème de la sélection** des unités lexicales complexes* (AR_0015)

2.2.3.1.2 Les bivalents

Le TSA est appliqué aux collocations *problème de l'évaluation* et *problème du choix*.

Problème de l'évaluation

SL – noyau 1 : *problème de l'évaluation* → question à résoudre (par une solution) → Donc appréciation, estimation, expertise, comparaison... Donc correction, rectification...

SL – noyau 2 : *problème de l'évaluation* → difficulté à expliquer/résoudre (en vue d'un résultat) → Donc appréciation, estimation, expertise, comparaison... Donc correction, rectification...

Problème du choix

SL – noyau 1 : *problème du choix* → question à résoudre (par une solution) → Donc décision entre plusieurs possibilités... Donc prendre de préférence...

SL – noyau 2 : *problème du choix* → difficulté à expliquer/résoudre (en vue d'un résultat) → Donc décision entre plusieurs possibilités... Donc prendre de préférence...

L'application du TSA montre que l'adjonction du collocatif à la base forme avec les stéréotypes proposés ou envisageables de *problème(s)* un bloc d'argumentation interne (H31.1). Par conséquent, nous concluons pour ces deux collocations à la concrétisation d'un PA dans le discours (H31.2). Il s'agit de collocations argumentatives formant un bloc d'argumentation externe.

Pour statuer sur leur axiologisation intrinsèque (H31.3), nous appliquons le TEA :

Problème de l'évaluation → Donc défaillance du logiciel

Problème de l'évaluation → *Pourtant défaillance du logiciel

Problème de l'évaluation → *Donc efficacité du logiciel

Problème de l'évaluation → Pourtant efficacité du logiciel

Seuls les enchaînements normatif « défaillance du logiciel » et transgressif « efficacité du logiciel » sont sémantiquement acceptables, ce qui signifie que l'argumentation est intrinsèquement axiologisée négativement. De fait, rappelons que l'unité lexicale *problème* est un monovalent négatif, tandis que les collocatifs sont bivalents. Que la base ou le collocatif occupe la position de tête, le bloc d'argumentation déployé est de nature normative et l'ensemble de la collocation est axiologisé négativement. La base conserve donc son potentiel axiologique et contamine négativement la bivalence du collocatif.

Par conséquent, au niveau macro-sémantique, l'axiologisation négative des bivalents résultant de la monovalence négative de la base *problème*, nous statuons sur un mécanisme d'activation de l'axiologisation de la signification.

Au niveau discursif, l'étude des collocations fait apparaître l'existence de quatre mécanismes possibles, dans l'ordre, de renforcement (avec *épineux*), de neutralisation (avec *aborder*, *évoquer*, *faire apparaître*, *être*), d'affaiblissement (avec *laisser en suspens* et la juxtaposition d'une négation et d'*offrir*, *traiter*) et d'interversion (avec *différentes solutions*, *apporter une réponse*).

Renforcement

Ex : *l'épineux problème du choix* du type de projection (assimilation ou règles) pour une configuration spatiale donnée n'a pas été automatisé (AC_0114)

Neutralisation

Ex : nous abordons à cet endroit le **problème de l'évaluation** d'un concept appris seulement à partir d'exemples positifs (AC_0011).

Ex : nous évoquerons le **problème de l'évaluation** (AC_0089)

Ex : l'exemple fait apparaître le **problème du choix** du type de détermination (AC_0046)

Ex : la présence de plusieurs sens possibles pour un même exemple est également **source de problèmes** (AC_0085)

Affaiblissement

Ex : les résultats de DISC n'offrent pas de réponse spécifique à un **problème d'évaluation** particulier, ils sont plus un ensemble de spécifications détaillées documentées (AR_0054)

Interversion

Ex : *examiner différentes solutions algorithmiques apportées au **problème de l'évaluation** de la résolution de la référence* (AR_0053)

Ex : *elle permet d'apporter une réponse nouvelle au **problème du choix** délicat du nombre de classes* (AR_0011)

2.2.3.2 Le collocatif est un adjectif

2.2.3.2.1 Les monovalents positifs

Nous appliquons le TSA à l'ensemble des collocations suivantes : *problème crucial*, *problème essentiel*, *problème important*, *problème majeur*, *problème principal*, *problème pratique*, *problème technique* et *vrais problèmes*.

Problème crucial

Comme pour *source du problème*, nous avons vérifié que la collocation *problème crucial* n'était pas lexicalisée. Nous n'en trouvons aucune trace sous l'entrée *problème* ; en revanche, elle est citée en exemple sous l'entrée *crucial* dans le TLFi. Nous la conservons donc pour l'analyse. En outre, l'acception de *crucial* au sens de « fondamental, très important » découle directement de sa signification philosophique « qui permet de confirmer ou rejeter une hypothèse ». Dans la mesure où ces éléments de signification éclairent le caractère argumentatif de la collocation, nous choisissons de faire figurer explicitement les éléments de signification dans les déploiements stéréotypiques. Le premier bloc d'argumentation mobilise donc la SL de *crucial* au sens philosophique du mot.

SL – noyau 1 : *problème crucial* → question à résoudre (par une solution) → Donc confirmation ou rejet d'une hypothèse... Donc fondamental, décisif, très important...

SL – noyau 2 : *problème crucial* → difficulté à expliquer/résoudre (en vue d'un résultat) → Donc confirmation ou rejet d'une hypothèse... Donc fondamental, décisif, très important...

Problème essentiel

SL – noyau 1 : *problème essentiel* → question à résoudre (par une solution) → Donc le plus important, principal, fondamental, primordial, capital... Donc véritable...

SL – noyau 2 : *problème essentiel* → difficulté à expliquer/résoudre (en vue d'un résultat) → Donc le plus important, principal, fondamental, primordial, capital... Donc véritable...

Problème important

SL – noyau 1 : *problème important* → question à résoudre (par une solution) → Donc principal, fondamental, primordial, capital... Donc véritable...

SL – noyau 2 : *problème important* → difficulté à expliquer/résoudre (en vue d'un résultat) → Donc principal, fondamental, primordial, capital... Donc véritable...

Problème majeur

SL – noyau 1 : *problème majeur* → question à résoudre (par une solution) → Donc d'une grande importance, considérable, exceptionnel...

SL – noyau 2 : *problème majeur* → difficulté à expliquer/résoudre (en vue d'un résultat) → Donc d'une grande importance, considérable, exceptionnel...

Problème principal

SL – noyau 1 : *problème principal* → question à résoudre (par une solution) → Donc le plus important, capital, essentiel, fondamental, majeur, primordial...

SL – noyau 2 : *problème principal* → difficulté à expliquer/résoudre (en vue d'un résultat) → Donc le plus important, capital, essentiel, fondamental, majeur, primordial...

Problème pratique

SL – noyau 1 : *problème pratique* → question à résoudre (par une solution) → Donc intervention humaine... Donc efficacité, pragmatisme...

SL – noyau 2 : *problème pratique* → difficulté à expliquer/résoudre (en vue d'un résultat) → Donc intervention humaine... Donc efficacité, pragmatisme...

Problème technique

SL – noyau 1 : *problème technique* → question à résoudre (par une solution) → Donc procédé professionnel, savoir-faire... Donc habileté, maîtrise...

SL – noyau 2 : *problème technique* → difficulté à expliquer/résoudre (en vue d'un résultat) → Donc procédé professionnel, savoir-faire... Donc habileté, maîtrise...

Vrais problèmes

SL – noyau 1 : *vrais problèmes* → question à résoudre (par une solution) → Donc réel, véritable, authentique...

SL – noyau 2 : *vrais problèmes* → difficulté à expliquer/résoudre (en vue d'un résultat) → Donc réel, véritable, authentique...

L'application du TSA montre que l'adjonction du collocatif à la base *problème(s)* forme un déploiement argumentatif possible (H32.1). Par conséquent, nous concluons pour toutes ces collocations à un mécanisme de motivation argumentative (H32.2).

Pour statuer sur leur axiologisation intrinsèque (H32.3), nous appliquons le TEA :

Problème crucial → Donc défaillance du logiciel

Problème crucial → *Pourtant défaillance du logiciel

Problème crucial → *Donc efficacité du logiciel

Problème crucial → Pourtant efficacité du logiciel

Seuls les enchaînements normatif « défaillance du logiciel » et transgressif « efficacité du logiciel » sont sémantiquement acceptables, ce qui signifie que le déploiement argumentatif est intrinsèquement axiologisé négativement. L'unité lexicale *problème(s)* est un monovalent négatif, tandis que les collocatifs mobilisés sont, eux, monovalents positifs. Contrairement aux collocations construites avec des noms monovalents positifs, dont l'axiologisation diffère selon la position tête de la base ou du collocatif, dans le cas des adjectifs le déploiement des blocs d'argumentations dévoile un seul schéma. Que la base ou le collocatif occupe la position de tête, l'argumentation déployée est de nature normative et l'ensemble de la collocation est axiologisé négativement. La base conserve donc son potentiel axiologique.

Au niveau du mécanisme macro-sémantique d'axiologisation de la signification (H41), la contamination de la polarité positive du collocatif par celle négative de la base produit un mécanisme de renforcement des valeurs axiologiques négatives portées par *problème(s)*, et ce que le collocatif suive ou précède la base.

Au niveau discursif, l'étude des collocations fait apparaître la possibilité de cinq mécanismes, de renforcement (avec *obscurcir*, *très grand nombre*), d'activation (avec *demeurer*, *mettre de côté*, *soulever*, *poser*), de neutralisation (avec *devenir*, *constituer*), d'affaiblissement (avec *tenter d'ébaucher*, l'association de *sembler* et d'une négation, avec *le mérite de proposer*, *subsister néanmoins*) et d'interversion (avec *résoudre*).

Renforcement

Ex : *l'utilisation d'exemples ciblés a obscurci un **problème crucial*** (AR_0071)

Ex : *un très grand nombre des **vrais problèmes** de syntaxe [...] sont clairement posés et pour certains expliqués* (AC_0062)

Activation

Ex : *d'un point de vue théorique, 3 **problèmes essentiels** demeurent* (AC_0022)

Ex : *l'utilité des informations [...] soulève un **problème important** d'alimentation* (AR_0060)

Ex : *les dépendances à distance posent un **problème important** pour l'analyse* (AC_0031)

Ex : *ces outils nous semblent mettre de côté les **vrais problèmes** relatifs à la construction d'une ontologie* (AR_0068)

Neutralisation

Ex : *c'est en particulier les modèles de langage qui permettent de choisir entre plusieurs formes homophones d'un même mot, ce qui constitue un des **problèmes cruciaux** des SRAP appliqués à la langue française* (AC_0018)

Ex : *ceci devient un **problème majeur** dans des situations où le corpus est en constante modification* (AC_0024)

Affaiblissement

Ex : *il semble que ce ne soit pas un **problème essentiel**, mais uniquement une tâche de plus à accomplir* (AR_0009)

Ex : *un **problème pratique** subsiste néanmoins concernant ce type d'étude portant sur la*

Ex : *nous tenterons ensuite d'ébaucher une réflexion sur le **problème technique** de la représentation formelle du discours* (AR_0032)

Ex : *la DRT a eu le mérite de proposer des premières solutions à un certain nombre de **problèmes techniques*** (AR_0019)

distance dans l'appariement (AR_0074)

Interversion

Ex : *nous aurons ainsi proposé une approche à l'analyse syntaxique [...] tout en résolvant certains des **problèmes importants** qu'elle pose* (AC_0120)

Ex : *plusieurs travaux ont été proposés pour résoudre les **problèmes pratiques** d'écriture et de mise à jour d'une LTAG* (AR_0064)

2.2.3.2.2 Les monovalents négatifs

Le TSA est appliqué aux trois collocations *problème complexe*, *problème difficile* et *problème épineux*.

Problème complexe

SL – noyau 1 : *problème complexe* → question à résoudre (par une solution) → Donc difficile, compliqué...

SL – noyau 2 : *problème complexe* → difficulté à expliquer/résoudre (en vue d'un résultat) → Donc difficile, compliqué...

Problème difficile

SL – noyau 1 : *problème difficile* → question à résoudre (par une solution) → Donc effort, peine... Donc laborieux, pénible...

SL – noyau 2 : *problème difficile* → difficulté à expliquer/résoudre (en vue d'un résultat) → Donc effort, peine... Donc laborieux, pénible...

Problème épineux

SL – noyau 1 : *problème épineux* → question à résoudre (par une solution) → Donc plein de difficultés... Donc délicat, embarrassant...

SL – noyau 2 : *problème épineux* → difficulté à expliquer/résoudre (en vue d'un résultat) → Donc plein de difficultés... Donc délicat, embarrassant...

L'application du TSA montre que l'adjonction du collocatif à la base forme avec les stéréotypes proposés ou envisageables de *problème(s)* un bloc d'argumentation interne (H31.1). Par conséquent, nous concluons pour ces collocations à la concrétisation de PA dans le discours (H31.2). Il s'agit de collocations argumentatives formant des blocs d'argumentations externes.

Pour statuer sur leur axiologisation intrinsèque (H31.3), nous appliquons le TEA :

Problème complexe → Donc défaillance du logiciel

Problème complexe → *Pourtant défaillance du logiciel

Problème complexe → *Donc efficacité du logiciel

Problème complexe → Pourtant efficacité du logiciel

Seuls les enchaînements normatif « défaillance du logiciel » et transgressif « efficacité du logiciel » sont sémantiquement acceptables, ce qui signifie que le PA est intrinsèquement axiologisé négativement. Ce paradigme de collocation n'offre qu'une construction pour laquelle la base est en position de tête. Les blocs d'argumentations déployés sont tous de nature normative et l'ensemble de la collocation est axiologisé négativement. La base conserve donc son potentiel axiologique.

Au niveau macro-sémantique, l'adjonction de deux polarités orientées négativement contribue à renforcer la valeur négative de la collocation toute entière. Nous statuons donc sur un mécanisme de renforcement de l'axiologisation négative.

Au niveau discursif, l'étude cotextuelle des collocations fait émerger deux mécanismes possibles, celui de l'activation (avec *rester*) et celui de la neutralisation (avec *représenter* et *être*).

Activation

Ex : *l'interprétation d'expressions référentielles reste un **problème difficile*** (AR_0003)

Neutralisation

Ex : *les dépendances entre les séquences représentent un **problème complexe*** (AR_0043)

Ex : *le **problème épineux** est donc de déterminer quels doivent être les effets fondamentaux des actes* (AR_0056)

2.2.3.2.3 Les bivalents

L'application du TSA concerne les collocations *problème classique*, *problème général*, *problème particulier* et *problème spécifique*.

Problème classique

SL – noyau 1 : *problème classique* → question à résoudre (par une solution) → *Donc habituel, courant... Donc banal, ordinaire...

SL – noyau 1 : *problème classique* → question à résoudre (par une solution) → Pourtant habituel, courant... Donc banal, ordinaire...

SL – noyau 2 : *problème classique* → difficulté à expliquer/résoudre (en vue d'un résultat) → *Donc habituel, courant... Donc banal, ordinaire...

SL – noyau 2 : *problème classique* → difficulté à expliquer/résoudre (en vue d'un résultat) → Pourtant habituel, courant... Donc banal, ordinaire...

Problème général

SL – noyau 1 : *problème général* → question à résoudre (par une solution) → *Donc commun à un ensemble de cas, universel... *Donc commun, courant, habituel...

SL – noyau 1 : *problème général* → question à résoudre (par une solution) → Pourtant commun à un ensemble de cas, universel... Pourtant commun, courant, habituel...

SL – noyau 2 : *problème général* → difficulté à expliquer/résoudre (en vue d'un résultat) → *Donc commun à un ensemble de cas, universel... *Donc commun, courant, habituel...

SL – noyau 2 : *problème général* → difficulté à expliquer/résoudre (en vue d'un résultat) → Pourtant commun à un ensemble de cas, universel... Pourtant commun, courant, habituel...

Problème particulier

SL – noyau 1 : *problème particulier* → question à résoudre (par une solution) → Donc limité à un élément, une partie... Donc singulier, spécifique...

SL – noyau 2 : *problème particulier* → difficulté à expliquer/résoudre (en vue d'un résultat) → Donc limité à un élément, une partie... Donc singulier, spécifique...

Problème spécifique

SL – noyau 1 : *problème spécifique* → question à résoudre (par une solution) → Donc caractéristique originale et exclusive... Donc particulier, spécial...

SL – noyau 2 : *problème spécifique* → difficulté à expliquer/résoudre (en vue d'un résultat) → Donc caractéristique originale et exclusive... Donc particulier, spécial...

L'application du TSA montre que l'adjonction du collocatif à la base *problème(s)* un déploiement argumentatif possible (H32.1). Par conséquent, nous concluons pour ces collocations à un mécanisme de motivation (H32.2).

Pour statuer sur leur axiologisation intrinsèque (H32.3), rappelons que l'unité lexicale *problème(s)* est un monovalent négatif, tandis que les collocatifs mobilisés sont bivalents. L'aspect transgressif ou normatif du déploiement des blocs d'argumentations permet de dégager deux cas de figure.

Nous appliquons le TEA au bloc transgressif :

Problème classique → *Donc défaillance du logiciel

Problème classique → Pourtant défaillance du logiciel

Problème classique → Donc efficacité du logiciel

Problème classique → *Pourtant efficacité du logiciel

Seuls les enchaînements transgressif « défaillance du logiciel » et normatif « efficacité du logiciel » sont sémantiquement acceptables, ce qui signifie que le déploiement argumentatif est intrinsèquement axiologisé positivement. La base conserve donc son potentiel axiologique négatif, mais le déploiement transgressif de l'argumentation confère au collocatif une polarité positive.

Nous appliquons le TEA au bloc normatif :

Problème particulier → Donc défaillance du logiciel

Problème particulier → *Pourtant défaillance du logiciel

Problème particulier → *Donc efficacité du logiciel

Problème particulier → Pourtant efficacité du logiciel

Seuls les enchaînements normatif « défaillance du logiciel » et transgressif « efficacité du logiciel » sont sémantiquement acceptables, ce qui signifie que l'argumentation est intrinsèquement axiologisé négativement. L'argumentation déployée est de nature normative. La base conserve donc son potentiel axiologique et contamine négativement la bivalence du collocatif.

Au niveau macro-sémantique, dans le premier cas nous statuons sur un mécanisme d'affaiblissement (voire d'interversion) des valeurs négatives portées par la base *problème(s)*, tandis que dans le deuxième cas, l'axiologisation négative des bivalents résultant de la monovalence négative de la base *problème*, nous statuons sur un mécanisme d'activation de l'axiologisation de la signification.

Au niveau discursif, l'étude cotextuelle des collocations fait émerger quatre mécanismes possibles, de renforcement (avec *évidemment*), de neutralisation (avec *noter*, *toucher*, *consacrer*, *parler*, *se concentrer*), d'affaiblissement (avec l'association d'une négation et de *poser*, *proposer un contournement*) et d'interversion (avec *répertoire de solutions trouvées*).

Renforcement

Ex : *c'est évidemment le **problème général** de la composition auquel tout modèle sémantique doit faire face* (AR_0018)

Neutralisation

Ex : *on peut cependant noter ici le **problème classique** d'évaluation de la classification non supervisée* (AC_0057)

Ex : *le second point, issu du même commentaire, touche au **problème classique** de "l'homme mort"* (AR_0022)

Ex : *on va parler de **problèmes particuliers** propres à la langue, sans se rendre compte que ces problèmes sont en fait parfois générés par les concepteurs de systèmes* (AR_0030)

Ex : *la section 3 est consacrée au **problème particulier** de la spatialité* (AR_0025)

Ex : *en ne se concentrant que sur des **problèmes spécifiques*** (AR_0040)

Affaiblissement

Ex : *le même cardinal (trois) ne pose pas de **problème particulier*** (AC_0107)

Interversion

Ex : *un répertoire de solutions déjà trouvées en réponse à des **problèmes spécifiques** (de terminologie, de phraséologie)* (AC_0008)

2.2.3.2.4 Les neutres

L'application du TSA concerne les collocations *problème sémantique* et *problèmes linguistiques*.

Problème sémantique

SL – noyau 1 : *problème sémantique* → question à résoudre (par une solution) → *Donc relatif au sens...

SL – noyau 1 : *problème sémantique* → question à résoudre (par une solution) → *Pourtant relatif au sens...

SL – noyau 2 : *problème sémantique* → difficulté à expliquer/résoudre (en vue d'un résultat) → *Donc relatif au sens...

SL – noyau 2 : *problème sémantique* → difficulté à expliquer/résoudre (en vue d'un résultat) → *Pourtant relatif au sens...

Problèmes linguistiques

SL – noyau 1 : *problèmes linguistiques* → question à résoudre (par une solution) → *Donc relatif au langage...

SL – noyau 1 : *problèmes linguistiques* → question à résoudre (par une solution) → *Pourtant relatif au langage ...

SL – noyau 2 : *problèmes linguistiques* → difficulté à expliquer/résoudre (en vue d'un résultat) → *Donc relatif au langage ...

SL – noyau 2 : *problèmes linguistiques* → difficulté à expliquer/résoudre (en vue d'un résultat) → *Pourtant relatif au langage ...

Pour *problème sémantique* et *problèmes linguistiques*, l'application du TSA montre que l'adjonction du collocatif à la base ne forme pas avec les stéréotypes proposés ou envisageables de *problème(s)* un bloc d'argumentation interne, ni un déploiement argumentatif possible (H33.1). Par conséquent, ces collocations sont arbitraires (H33.2).

Nous appliquons donc le TEA pour savoir si la base conserve son potentiel axiologique négatif (H33.3) :

Problème sémantique → Donc défaillance du logiciel

Problème sémantique → *Pourtant défaillance du logiciel

Problème sémantique → *Donc efficacité du logiciel

Problème sémantique → Pourtant efficacité du logiciel

Seuls les enchaînements normatif « défaillance du logiciel » et transgressif « efficacité du logiciel » sont sémantiquement acceptables, ce qui signifie que la collocation est intrinsèquement axiologisée négativement, sans être intrinsèquement argumentative, mais tout en conservant son potentiel axiologique. L'orientation axiologique négative de la base *problème* contraint l'instanciation négative de l'ensemble de la collocation, sans que nous ayons la preuve qu'elle « contamine » la neutralité du collocatif.

Au niveau discursif, l'observation co- et contextuelle des collocations fait émerger trois mécanismes possibles, le renforcement (avec l'association de *beaucoup* et de *soulever*), de l'affaiblissement (avec *identifier*, *s'intéresser*), et d'interversion (avec l'association de *rencontrer* et *traiter*).

Renforcement

Ex : *la lecture du document d'accompagnement montre que beaucoup de **problèmes linguistiques** ont été soulevés durant la phase proprement dite d'évaluation* (AR_0013)

Affaiblissement

Ex : *nous avons identifié des **problèmes linguistiques** auxquels nous devons nous adresser pour améliorer la qualité de notre traduction* (AC_0067)

Ex : *nous nous intéressons ensuite aux **problèmes linguistiques** à résoudre pour le français au cours des différentes phases de la reconnaissance des noms propres* (AR_0036)

Interversion

Ex : *les **problèmes linguistiques** rencontrés ont été traités plus particulièrement le jour de la réunion finale d'évaluation* (AR_0013)

2.3 Synthèse sur les monovalents négatifs

De même que pour les monovalents positifs, cette synthèse repose sur les analyses sémantico-discursives menées, concernant le binôme monovalent négatif *erreur(s)* – *problème(s)*. Ces deux bases rassemblent à elles deux 48 collocations différentes et 418 occurrences.

Les monovalents positifs ont généré quelques collocations argumentatives, parmi lesquelles *diagnostic d'erreur*, *réponse au problème*, *problème d'évaluation*, *problème complexe*, etc. Pour toutes ces collocations, l'application du TSA a montré que l'adjonction du collocatif à la base forme avec les stéréotypes proposés ou envisageables de *erreur(s)* ou *problème(s)* un bloc d'argumentation interne (H31.1). Aussi, toutes ces collocations correspondent à la concrétisation d'un PA dans le discours (H31.2) ; elles sont prévisibles, « calculables » sur la base des PA existant pour chaque mot. Nous rappelons que de notre point de vue, ces collocations ne formant pas un bloc d'argumentation externe au sens où aucun connecteur argumentatif n'articule le collocatif à la base, elles correspondent à la réalisation de surface d'un bloc d'argumentation externe sous-jacent.

Conformément aux postulats théoriques de la SPA-ALD, de par le lien argumentatif posé, les PA sont par essence évaluatifs. Le TEA a montré que toutes les collocations argumentatives relevées sont axiologisées soit négativement comme pour *problème difficile*, soit positivement comme pour *résolution du problème*.

Le PA est axiologisé négativement lorsque le collocatif nominal ou adjectival, postposé par rapport à la base, est négatif ou bivalent. Selon les cas, le mécanisme macro-sémantique varie en fonction de la nature axiologique du collocatif :

- quand le collocatif est négatif, l'adjonction de deux polarités négatives crée un mécanisme de renforcement (H41), comme pour *problème épineux* ;
- quand le collocatif est bivalent, l'adjonction de la polarité négative de la base semble contraindre l'orientation négative du collocatif créant un mécanisme d'activation (H41), comme pour *diagnostic d'erreur* ; *problème du choix*.

Le PA est axiologisé positivement lorsque le collocatif nominal, antéposé par rapport à la base, est positif. Dans ce cas, l'antéposition⁷ du collocatif sur la base influe sur l'axiologisation négative de la base, créant un mécanisme d'interversion (H41), comme pour *solution du problème* par exemple.

Quatre mécanismes discursifs ont été rencontrés : le renforcement (**Ex** : *risque de se tromper* dans le diagnostic d'erreur ; *excellents résultats peuvent permettre* de résoudre les problèmes), la neutralisation (**Ex** : *établir* un diagnostic d'erreur), l'affaiblissement (**Ex** : *le résultat n'offre pas de réponse* au problème d'évaluation) et l'interversion (**Ex** : *permet d'apporter une réponse nouvelle* au problème du choix).

Les monovalents négatifs ont également généré beaucoup de collocations argumentativement motivées, parmi lesquelles *réduction des erreurs*, *erreur fréquente*, *problème de cohérence*, *problème crucial*, *problème spécifique*, etc. Pour toutes les collocations concernées, l'application du TSA a montré que l'adjonction du collocatif à la base *erreur(s)* ou *problème(s)* forme un déploiement argumentatif possible (H32.1). Aussi, toutes ces collocations correspondent à un mécanisme de motivation argumentative (H32.2). Elles entretiennent une relation argumentative avec la base. Le mécanisme de motivation argumentative revêt un caractère moins prédictible que celui mobilisé par les collocations argumentatives, sans que l'émergence de régularités en fonction de la monovalence négative de la base soit compromise. De même que pour les collocations argumentatives, nous trouvons plus juste de parler de réalisation de surface d'un bloc d'argumentation externe sous-jacent.

Conformément aux postulats théoriques de la SPA-ALD, de par le lien argumentatif déployé, ces collocations comportent nécessairement une évaluation, orientée vers l'un ou l'autre des pôles axiologiques, positif ou négatif (H32.3). Le TEA a montré que pour tous les cas de motivation argumentative, l'orientation axiologique des collocations est soit négative comme pour *problème particulier*, soit positive comme pour *diminution des erreurs*.

La collocation est axiologisée négativement lorsque le collocatif, qu'il s'agisse d'un nom ou d'un adjectif postposé par rapport à la base, est négatif, bivalent, neutre ou positif. Toutefois, le mécanisme macro-sémantique varie en fonction de la nature axiologique du collocatif :

- quand le collocatif est négatif, l'adjonction de deux polarités négatives crée un mécanisme de renforcement (H41), comme pour *risque d'erreur* ;
- quand le collocatif est bivalent, l'adjonction de la polarité négative de la base semble contraindre l'orientation négative du collocatif créant un mécanisme d'activation (H41), comme pour *problème particulier* ;
- quand le collocatif est neutre, l'adjonction de la polarité négative de la base semble « contaminer » négativement la neutralité du collocatif, créant un mécanisme d'activation (H41), comme pour *erreur d'accord* ;

⁷ C'est bien l'antéposition du collocatif à valeur positive qui inverse la polarité de la collocation, car en cas de postposition, la collocation conserve son potentiel axiologique négatif. **Ex** : *problème de solution*.

- quand le collocatif est positif, la postposition de la polarité positive portée par le collocatif influe sur la polarité négative de la base, dans un mécanisme de renforcement (H41), comme pour *problème important*.

La collocation est axiologisée positivement lorsque le collocatif, qu'il s'agisse d'un nom ou d'un adjectif, est positif, bivalent ou négatif. Les régularités observées sont les suivantes :

- quand le collocatif est positif, son antéposition par rapport à la base négative crée un mécanisme d'interversion (H41), comme pour *analyse d'erreur* ;
- quand le collocatif est bivalent et que le déploiement argumentatif est transgressif, l'adjonction de la polarité positive de la base négative, crée un mécanisme d'interversion (H41), comme pour *problème classique* ;
- quand le collocatif est négatif, son antéposition par rapport à la base crée un mécanisme d'interversion (H41), comme pour *réduction d'erreur*.

Cinq mécanismes discursifs ont été rencontrés : le renforcement (**Ex** : *se heurter systématiquement* à un problème de cohérence), l'activation (**Ex** : *permettre* de réduire les erreurs), la neutralisation (**Ex** : *répartir* les erreurs d'accord), l'affaiblissement (**Ex** : *faible* risque d'erreur) et l'interversion (**Ex** : *exposer une solution flexible et robuste* au problème de la sélection).

Enfin, nous avons trouvé quelques collocations arbitraires formées sur les bases *erreur(s)* et *problème(s)* : *repérage d'erreur*, *problème sémantique*, etc. La collocation arbitraire correspond au cas où la collocation repérée ne correspond ni à la concrétisation d'un PA dans le discours, ni à un cas de motivation argumentative (H33.1). Dans ce cas, la collocation est qualifiée d'arbitraire (H33.2), au sens où elle ne peut être sémantiquement prédictible ; elle n'a de réalité que discursive.

Par l'absence de lien argumentatif, la collocation arbitraire ne devrait pas être intrinsèquement axiologisée, ce que nous discutons. Le TEA a montré que les collocations étaient axiologisées soit négativement, soit positivement (H33.3). La collocation est axiologisée négativement lorsque le collocatif nominal ou adjectival, antéposé ou postposé par rapport à la base, est neutre, comme pour *pourcentage d'erreur* et *problèmes linguistiques* par exemples. Dans ce cas, il s'agit d'un mécanisme macro-sémantique d'activation (H41). La collocation est axiologisée positivement lorsque le collocatif nominal est positif, antéposé par rapport à la base, créant ainsi un mécanisme d'interversion (H41), comme pour *repérage d'erreur*.

Quatre mécanismes discursifs ont été rencontrés : le renforcement (**Ex** : *soulever beaucoup* de problèmes linguistiques), la neutralisation (**Ex** : *distinguer* les erreurs), l'affaiblissement (**Ex** : problèmes linguistiques à *résoudre*) et l'interversion (**Ex** : les problèmes linguistiques rencontrés *ont été traités*).

En confrontant les mécanismes obtenus pour les trois types de collocation, aucune tendance ne se dégage pour statuer sur l'hypothèse H42 relative à la prépondérance des mécanismes discursifs de renforcement et d'activation dans le discours scientifique spécialisé.

3 Les bivalents

3.1 Évaluer – Évaluation(s)

3.1.1 Dépouillement des données

3.1.1.1 Fréquences et nettoyage

Le corpus TAL(N) comporte 809 occurrences de *évaluation* au singulier et 63 occurrences du mot au pluriel. Le verbe *évaluer* apparaît 215 fois à l’infinitif et 193 fois sous forme fléchie (figure entre parenthèse les fréquences respectives des formes fléchies) : évaluant (4), évalue (29), évalué (59), évaluée (44), évaluées (11), évaluent (7), évaluerons (4), évalués (18), évaluons (15).

Selon le critère statistique, nous éliminons les candidats trop peu fréquents. Nous retirons donc tous les apax, les candidats de fréquence deux et les cooccurrents n’apparaissant pas dans trois articles distincts.

Ex : *impossible d’évaluer* (1), *évalue une quantité* (1), *conditions d’évaluation* (1), *contexte d’évaluation* (1), *typologie d’évaluation* (1).

Selon le critère de pertinence, nous laissons de côté toutes les collocations grammaticales et autres constructions dont le collocatif ne présente pas de contenu sémantique.

Ex : *évaluation(s) de/des*, *évaluer le/la/les*.

Selon le critère lexical, aucune lexie construite sur *évaluation(s)* ou *évaluer* n’a été trouvée en corpus.

En revanche, selon le critère lexicographique, le corpus TAL(N) comporte trois termes complexes construits sur *évaluation* : *campagne d’évaluation* (58), *conférence d’évaluation* (33), *corpus d’évaluation* (4). Toutefois, ces termes ne sont pas mentionnés ni dans le GDT ni dans Eurodicautom, ils ont été validés par des experts du domaine. En effet, *campagne d’évaluation* est systématiquement suivie d’un acronyme ou d’un nom propre désignant un programme (AMARYLLIS, ARPA ATIS, GRACE, MULTITAG, MUC) et *conférence d’évaluation* n’est associée qu’à un seul sigle, celui de TREC. À propos de *corpus d’évaluation*, nous avons lancé plusieurs requêtes dans les dictionnaires pour savoir si quelques termes complexes construits sur corpus étaient enregistrés. Or aucun des termes suivants ne figure dans les dictionnaires : *corpus d’apprentissage*, *corpus de test*, *corpus comparable*, *alignement de corpus*, *corpus attesté*, *corpus électronique*, *corpus parallèle*, *corpus ouvert/fermé*. Il semblerait que le développement récent de ces techniques soit à l’origine de leur absence dans les dictionnaires.

La collocation *évaluation automatique* (10) correspond à un terme dans le domaine de la gestion traduit par *built-in evaluation*, mais nous n’avons trouvé aucune occurrence de ce terme complexe dans le GDR et Eurodicautom, réservé aux domaines informatique et/ou linguistique.

3.1.1.2 Liste des collocations lexicales

Cette sous-section présente les cinq tableaux récapitulatifs des collocations lexicales construites sur *évaluation(s)* et *évaluer*. Les trois premiers tableaux correspondent aux contextes droit et gauche du nom et les deux suivants, aux contextes droit et gauche du verbe.

COLLOCATIONS LEXICALES	FREQ.	AUTRES FORMES	FREQ.	TOTAL
Évaluation automatique	8	-Évaluations automatiques -Évaluer de manière automatique	1 1	10
Évaluation comparative	11	-Évaluations comparatives	3	14
Évaluation de la cohésion	1	-Évaluer la cohésion -Évaluer [...] de cohésion	3 5	9
Évaluation de la grammaire	1	-É. de grammaire -Évaluer [...] grammaire(s)	1 2	4
Évaluation de la pertinence	2	-É. [...] la pertinence -Évaluer la pertinence	1 5	8
Évaluation de la qualité	7	-É. de qualité -Évaluer la qualité	2 6	15
Évaluation des algorithmes	3	-É. de l'algorithme -É. de nos algorithmes -Évaluer [...] algorithme(s)	1 1 4	9
Évaluation des méthodes	3	-É. de [...] méthode(s) -Évaluer une méthode -La méthode doit être É.	3 1 1	8
Évaluation des performances	2	-É. [...] des performances -Évaluer les performances -É. [...] performance(s) -Les performances [...] être É	2 13 4 1	22
Évaluation des réponses	3	-É. de la réponse -Évaluer [...] réponse(s)	2 5	10
Évaluation des résultats	9	-É. [...] résultats -Évaluer les résultats -É. [...] résultat(s) -Les résultats [...] être É.	3 11 5 1	33
Évaluation des systèmes	17	-É. [...] système(s) -É. du système -É. d'un système -Évaluer les systèmes -É. [...] système(s) -Le système [...] être É.	12 8 6 7 7 5	43
Évaluation du modèle	4	-É. [...] modèle(s) -Le modèle [...] être É.	5 2	11
Évaluation du processus	2	-Évaluer le processus	1	3
Évaluation empirique	4	-	-	4
Évaluation fine	3	-É. [...] fine -Évaluer plus finement	2 1	6
Évaluation globale	3	-É. [...] globale	1	5
Évaluation manuelle	4	-É. est faite manuellement -Évaluations manuelles -Évaluer manuellement	1 1 2	9
Évaluation objective	6	-Évaluations objectives	2	8
Évaluation qualitative	3	-Évaluations [...] qualitatives	1	4
Évaluation quantitative	5	-	-	5

Tableau 21 : Collocations lexicales contexte droit de évaluation(s)

COLLOCATIONS LEXICALES	FREQ.	AUTRES FORMES	FREQ.	TOTAL
Évaluations expérimentales	1	-Évaluer expérimentalement -Évaluer [...] expérimentale	1 1	3

Tableau 22 : Collocations lexicales contexte droit de évaluations

COLLOCATIONS LEXICALES	FREQ.	AUTRES FORMES	FREQ.	TOTAL
Coût de l'évaluation	2	-Coût des évaluations -Coût [...] des évaluations	1 1	4
Critère(s) d'évaluation	10	-Critères [...] d'E.	1	11
Mesure(s) d'évaluation	10	-	-	10
Méthodologie(s) d'évaluation	13	-	-	13
Outils d'évaluation	8	-	-	8
Paradigme(s) d'évaluation	22	-	-	22
Phase(s) d'évaluation	8	-Phase d'évaluations	1	9
Procédure(s) d'évaluation	3	-	-	3
Processus d'évaluation	5	-	-	5
Programme(s) d'évaluation	4	-	-	4
Protocole(s) d'évaluation	26	-	-	26
Technique(s) d'évaluation	3	-	-	3
Travail d'évaluation	3	-	-	3

Tableau 23 : Collocations lexicales contexte gauche de évaluation(s)

COLLOCATIONS LEXICALES	FREQ.	AUTRES FORMES	FREQ.	TOTAL
Difficile à évaluer	6	-Difficile d'E.	2	8
Intéressant d'évaluer	3	-	-	3

Tableau 24 : Collocations lexicales contexte gauche de évaluer

3.1.2 Signification lexicale de évaluer – évaluation

3.1.2.1 Signification lexicale de évaluer

Le LL propose une entrée pour *évaluer* sous laquelle sont regroupés évaluable, évaluation, évaluatif-ive, évaluateur, réévaluer, réévaluation, sous-évaluer, sous-évaluation, surévaluer, surévaluation (cf. Annexe 8). Une seule définition est proposée pour ce verbe transitif.

Définition : En déterminer plus ou moins approximativement la valeur, l'importance.

→ estimer (*évaluer à mille francs par jour le chiffre d'affaires moyen ; évaluer exactement la distance*)

Le GR propose une entrée pour *évaluer* décomposée en deux définitions, sous laquelle figure également en sous-entrée le participe passé/adjectif évalué (cf. Annexe 8). À noter que dans la première définition, bien que ce ne soit pas explicitement indiqué, figure une acception référant au domaine de l'expertise (estimer, priser, expertiser → *faire évaluer un tableau*), que nous laissons de côté pour la construction de la SL.

Définition 1 : Déterminer (une quantité) par le calcul (sans recourir à la mesure directe).

→ calculer, cuber, jauger, commensurable, immensurable (*évaluer un volume ; évaluer le débit d'une rivière ; possible-impossible à évaluer*)

Définition 2 : Fixer approximativement.

→ apprécier, estimer, juger (*évaluer une distance*)

→ apprécier, déterminer, juger (*évaluer les qualités de qqn.*)

Aucune définition n'est associée au participe passé/adjectif, et les deux exemples mentionnés reprennent les deux définitions associées au verbe, *évaluer une distance à vue d'œil* et *foule évaluée à deux milles personnes environ*. Aussi, nous ne retenons aucune définition.

Figure ci-dessous le tableau de synthèse de la signification lexicale du verbe.

ÉVALUER			
TRAITS NECESSAIRES DE CATEGORISATION (TNC)			
TYPE GRAMMATICAL		ZONE MODALE	ORIENTATION AXIOLOGIQUE
Verbe – Transitif		Axiologique	Bivalent
ENTITES SEMANTIQUES			ENTITES DISCURSIVES
NOYAU		STEREOTYPES	POSSIBLES ARGUMENTATIFS
1	Fixer plus ou moins approximativement la valeur, l'importance de qqch	Donc estimer, juger... Donc apprécier, déterminer...	E. donc estimer une distance E. donc déterminer un risque
2	Déterminer (une quantité) par le calcul (sans mesure directe)	Donc calculer, cuber, jauger... Donc commensurable...	E. donc calculer un volume E. donc comparer

Tableau 25 : Signification lexicale de évaluer

3.1.2.2 Signification lexicale de évaluation

Comme pour *amélioration*, *évaluation* dans le LL figure une sous-entrée de l'article *évaluer* (cf. Annexe 8). Ce sous-article ne propose pas de définition, au sens où on n'y trouve pas de périphrase synonymique du défini, mais un seul exemple construit – *Une rapide évaluation du montant des réparations fait prévoir trente mille franc de frais* –, synonyme de *estimation*.

Nous nous reportons donc à la définition proposée pour *évaluer*.

Définition : En déterminer plus ou moins approximativement la valeur, l'importance.

Le GR propose une entrée pour *évaluation* (cf. Annexe 8). L'article propose deux définitions, l'une au sens d'évaluer qui rassemble les deux noyaux mentionnés ci-dessus, l'autre qui envisage le résultat de l'évaluation.

Définition 1 : Action d'évaluer.

→ appréciation, détermination, estimation, expertise, comparaison

→ calcul, inventaire, majoration (*évaluation d'une fortune ; évaluation des marchandises ; évaluation d'une chose*)

→ approximation, échelle (*évaluation d'un débit ; moyen d'évaluation*)

Définition 2 : Valeur, quantité évaluée.

→ mesure, prix, valeur (*évaluation insuffisante/trop faible ; corriger/rectifier une évaluation*)

Les deux dictionnaires décrivent *évaluation* comme un nom féminin. Cette unité lexicale appartient à la zone modale de l'axiologique, puisque le mot *valeur* participe de sa définition (de son noyau), ce qui, par répercussion, fait de lui un bivalent, une évaluation peut être positive ou négative.

Figure ci-dessous le tableau de synthèse de la signification lexicale du nom.

ÉVALUATION(S)			
TRAITS NECESSAIRES DE CATEGORISATION (TNC)			
TYPE GRAMMATICAL		ZONE MODALE	ORIENTATION AXIOLOGIQUE
Nom – Féminin		Axiologique	Bivalent
ENTITES SEMANTIQUES			ENTITES DISCURSIVES
NOYAU		STEREOTYPES	POSSIBLES ARGUMENTATIFS
1	Processus : action de fixer plus ou moins approximativement la valeur, l'importance de qqch	Donc appréciation, estimation, jugement... Donc comparaison, expertise... Donc calcul, inventaire...	E. donc appréciation des qualités E. donc expertise d'un tableau
2	Résultat : valeur, importance fixée (quantité)	Donc correction, rectification... Donc renouvellement... Donc mesure, qualification...	E. donc insuffisance E. donc bonne qualification

Tableau 26 : Signification lexicale de évaluation

Le noyau 1 du verbe et du nom étant similaire, nous ne soumettons les collocations au test de substitution argumentatif qu'une seule fois en les regroupant sous le même appellatif, Nom/verbe – noyau 1.

L'observation contextuelle des occurrences des collocations nous permet de déterminer la SL mobilisée. Nous mentionnons les cas de doute et appliquons dans ces situations toutes les SL construites ci-dessus.

En outre, la présence de valeur dans la SL de évaluation et évaluer par sa polysémie nous semble propre à générer des difficultés dans l'attribution du statut argumentatif *vs* arbitraire des collocations. Aussi, nous proposons ci-dessous la SL du mot, construite selon le même principe que les précédentes.

VALEUR(S)			
TRAITS NECESSAIRES DE CATEGORISATION (TNC)			
TYPE GRAMMATICAL		ZONE MODALE	ORIENTATION AXIOLOGIQUE
Nom – Féminin		Axiologique	Bivalent
ENTITES SEMANTIQUES		ENTITES DISCURSIVES	
NOYAU		STEREOTYPES	POSSIBLES ARGUMENTATIFS
1	Caractère mesurable d'un objet susceptible d'être échangé, désiré	Donc apprécier, estimer... Donc surévaluer, sous-évaluer...	V. donc estimation d'un bijou
2	Mesure d'une grandeur, d'un nombre ou d'une quantité variable	Donc comparer, évaluer... Donc calculer, mesurer...	V. donc mesure absolue
3	Mesure conventionnelle d'un signe dans une série	Donc apprendre, connaître...	V. donc utilisation du roi de cœur
4	Qualité estimée par un jugement	Donc considérer, estimer... Donc conclure, penser, savoir...	V. donc conclusion mitigée

Tableau 27 : Signification lexicale de valeur(s)

3.1.3 Analyse des données

3.1.3.1 Le collocatif est un nom

3.1.3.1.1 Les monovalents positifs

L'application du TSA concerne trois collocations, *évaluation de la cohésion*, *évaluation de la pertinence* et *évaluation des réponses*.

Évaluation de la cohésion

Nom/verbe – noyau 1 : *évaluation de la cohésion* → fixer approximativement la valeur → *Donc cohérence, unité logique...

Nom/verbe – noyau 1 : *évaluation de la cohésion* → fixer approximativement la valeur → Pourtant cohérence, unité logique...

Évaluation de la pertinence

Nom/verbe – noyau 1 : *évaluation de la pertinence* → fixer approximativement la valeur → *Donc adapté... *Donc à-propos, juste...

Nom/verbe – noyau 1 : *évaluation de la pertinence* → fixer approximativement la valeur → Pourtant adapté... Pourtant à-propos, juste...

Évaluation des réponses

Nom – noyau 2 : *évaluation des réponses* → valeur fixée → Donc solution, explication apportée par le raisonnement...

Verbe – noyau 2 : *évaluer les réponses* → déterminer par le calcul → Donc solution, explication apportée par le raisonnement...

L'application du TSA montre que l'adjonction du collocatif à la base *évaluer* et/ou de *évaluation(s)* forme un déploiement argumentatif possible (H32.1), qu'il soit déployé de façon transgressive avec *cohésion* et *pertinence* ou de façon normative comme pour *réponses*. Par conséquent, nous concluons pour chacune de ces collocations à un mécanisme de motivation argumentative (H32.2).

Pour statuer sur l'axiologisation intrinsèque (H32.3), nous recourons au TEA :

Évaluer la cohésion → Donc améliorer l'analyse

Évaluer la cohésion → *Pourtant améliorer l'analyse

Évaluer la cohésion → Donc altérer l'analyse

Évaluer la cohésion → *Pourtant altérer l'analyse

Évaluation de la cohésion → Donc amélioration de l'analyse

Évaluation de la cohésion → *Pourtant amélioration de l'analyse

Évaluation de la cohésion → Donc altération de l'analyse

Évaluation de la cohésion → *Pourtant altération de l'analyse

Dans cette série de données, la base occupe systématiquement la position tête de la collocation, nous n'avons donc pas de variation possible de l'axiologisation intrinsèque en fonction de la position du collocatif. Nous remarquons également que la répartition transgressive *vs* normative des déploiements argumentatifs est indexée sur le changement de noyau de la SL. Pour autant, l'application du TEA montre que tous les enchaînements normatifs « améliorer » et « altérer » sont sémantiquement acceptables, ce qui signifie que le déploiement argumentatif conserve un potentiel axiologique bivalent, en l'absence de tout élément co- ou contextuel. Les deux enchaînements transgressifs sont asémantiques, compte tenu de l'orientation potentielle positive et négative possible de chaque collocation.

Au niveau des mécanismes macro-sémantiques d'axiologisation de la signification (H41), la polarité positive du collocatif ne contamine pas la bivalence de la base, ce qui nous permet de statuer sur un mécanisme de neutralisation des valeurs axiologiques bivalentes portées par *évaluer* et *évaluation(s)*. Nous attribuons ce mécanisme à toutes les collocations ci-dessus analysées.

Au niveau discursif, l'observation co- et contextuelle des collocations fait apparaître l'existence de trois mécanismes : le renforcement (avec l'association de *permettre* et de *pistes fiables*), l'activation (avec *permettre*), la neutralisation (avec *chercher*, *réaliser*). Dans les cas de renforcement et d'activation, l'axiologisation est toujours positive, en revanche, dans le cas majoritairement répandu de la neutralisation, seule une étude contextuelle large (sur l'ensemble de l'article) permet de statuer sur une orientation positive (la plupart du temps) ou négative, puisque l'évaluation de tel ou tel outil représente généralement l'enjeu de l'article proposé.

Renforcement

Ex : *certaines indices semblent cependant intervenir très fortement et permettent d'envisager des pistes fiables pour une **évaluation de la pertinence*** (AC_0093)

Activation

Ex : *nous nous sommes ainsi dotées d'une mesure permettant d'**évaluer la cohésion** d'un GP pour un verbe donné* (AC_0059)

Neutralisation

Ex : *la partie du texte dont on cherche à **évaluer la cohésion*** (AC_0108).

Ex : *elle définit un champ d'investigation de taille réduite pour **évaluer la pertinence** d'une approche* (AR_0052)

Ex : *nous avons réalisé une **évaluation des réponses** que nous trouvons sur le corpus de la conférence TREC 10* (AC_0076)

3.1.3.1.2 Le monovalent négatif

Le TSA est appliqué à la collocation *coût de l'évaluation*.

Coût de l'évaluation

Dans cette collocation, *évaluation(s)* peut référer soit à l'aspect processus, soit à l'aspect résultat du noyau ; c'est pourquoi nous faisons figurer les deux déploiements argumentatifs possibles.

Nom – noyau 1 : *coût de l'évaluation* → fixer approximativement la valeur → Donc conséquence négative, effets supportés...

Nom – noyau 2 : *coût de l'évaluation* → valeur fixée → Donc conséquence négative, effets supportés...

L'application du TSA montre que l'adjonction du collocatif à la base *évaluation* forme un déploiement argumentatif possible (H32.1). Par conséquent, nous concluons pour *coût de l'évaluation* à un mécanisme de motivation argumentative (H32.2).

Nous utilisons le TEA pour statuer sur son axiologisation intrinsèque (H32.3) :

Coût de l'évaluation → Donc temps de travail élevé

Coût de l'évaluation → *Pourtant temps de travail élevé

Coût de l'évaluation → *Donc temps de travail réduit

Coût de l'évaluation → Pourtant temps de travail réduit

Dans ce cas, le collocatif est en position tête et le bloc d'argumentation est normatif. Seuls les enchaînements normatif avec « travail élevé » et transgressif avec « travail réduit » sont sémantiquement acceptables, ce qui signifie que le déploiement argumentatif est intrinsèquement axiologisé négativement.

Au niveau des mécanismes macro-sémantiques d'axiologisation de la signification (H41), la polarité négative du collocatif contamine donc la bivalence de la base, ce qui nous permet de statuer sur un mécanisme d'activation des valeurs axiologiques bivalentes portées par *évaluation(s)*.

Au niveau discursif, l'observation cotextuelle de la collocation fait apparaître un seul mécanisme d'interversion, avec la juxtaposition de *nécessité* et *réduire*.

Interversion

Ex : *le bilan de ces deux expériences nous a amené à considérer la nécessité de réduire le coût des évaluations* (AR_0054)

3.1.3.1.3 Les bivalents

Les collocations construites avec des collocatifs bivalents sont particulièrement nombreuses. L'application du TSA concerne les collocations *évaluation de la qualité*, *évaluation des performances*, *évaluation des résultats*, *évaluation du modèle*, *évaluation du processus*, *critère(s) d'évaluation*, *méthodologie(s) d'évaluation*, *paradigme(s) d'évaluation*, *procédure(s) d'évaluation*, *processus d'évaluation*, *programme(s) d'évaluation*, *protocole(s) d'évaluation*, *technique(s) d'évaluation* et *travail d'évaluation*.

Évaluation de la qualité

Nom/verbe – noyau 1 : *évaluation de la qualité* → fixer approximativement la valeur → Donc caractéristique, valeur...

Évaluation des performances

Dans cette collocation, *performances* est employé uniquement en qualité de bivalent, mais *évaluation* ne réfère qu'à l'aspect processus correspondant au noyau 1.

Nom/verbe – noyau 1 : *évaluation des performances* → fixer approximativement la valeur → Donc résultats d'une épreuve, d'un test...

Évaluation des résultats

Nom/verbe – noyau 1 : *évaluation des résultats* → fixer approximativement la valeur → Donc effets, conséquences...

Évaluation du modèle

Dans le cas de cette collocation, *évaluation* est employé à quatre reprises précédé de *auto-*, ce qui ne semble pas remettre en cause la SL de *évaluation*. Aussi, nous déployons les blocs argumentatifs sur la même SL.

Nom/verbe – noyau 1 : *évaluation du modèle* → fixer approximativement la valeur → Donc processus, système formalisé, scientifiquement testé...

Nom – noyau 2 : *évaluation du modèle* → valeur fixée → Donc processus, système formalisé, scientifiquement testé...

Verbe – noyau 2 : *évaluer les modèles* → déterminer par le calcul → Donc processus, système formalisé, scientifiquement testé...

Évaluation du processus

Nom – noyau 2 : *évaluation du processus* → valeur fixée → Donc procédure aboutissant à un résultat...

Verbe – noyau 2 : *évaluer le processus* → déterminer par le calcul → Donc procédure aboutissant à un résultat...

Critère(s) d'évaluation

Nom/verbe – noyau 1 : *évaluation des résultats* → fixer approximativement la valeur → Donc éléments servant de base à un jugement d'appréciation...

Méthodologie(s) d'évaluation

Nom – noyau 1 : *méthodologie(s) d'évaluation* → fixer approximativement la valeur → Donc règles, démarches adoptées pour conduire une recherche...

Nom – noyau 2 : *méthodologie(s) d'évaluation* → valeur fixée → Donc règles, démarches adoptées pour conduire une recherche...

Paradigme(s) d'évaluation

Nom – noyau 1 : *paradigme(s) d'évaluation* → fixer approximativement la valeur → Donc modèle qui fonde les types d'explications et types de faits dans une science donnée...

Nom – noyau 2 : *paradigme(s) d'évaluation* → valeur fixée → Donc modèle qui fonde les types d'explications et types de faits dans une science donnée...

Procédure(s) d'évaluation

Nom – noyau 1 : *procédure(s) d'évaluation* → fixer approximativement la valeur → Donc méthode, moyen, processus utilisés dans la conduite d'une opération technique ou scientifique...

Nom – noyau 2 : *procédure(s) d'évaluation* → valeur fixée → Donc méthode, moyen, processus utilisés dans la conduite d'une opération technique ou scientifique...

Processus d'évaluation

Nom – noyau 1 : *processus d'évaluation* → fixer approximativement la valeur → Donc procédure aboutissant à un résultat...

Nom – noyau 2 : *processus d'évaluation* → valeur fixée → Donc procédure aboutissant à un résultat...

Programme(s) d'évaluation

Nom – noyau 1 : *programme(s) d'évaluation* → fixer approximativement la valeur → Donc succession d'actions aboutissant à un résultat...

Nom – noyau 2 : *programme(s) d'évaluation* → valeur fixée → Donc succession d'actions aboutissant à un résultat...

Protocole(s) d'évaluation

Nom – noyau 1 : *protocole(s) d'évaluation* → fixer approximativement la valeur → Donc succession d'opérations à respecter pour exécuter une opération, réaliser une expérience...

Nom – noyau 2 : *protocole(s) d'évaluation* → valeur fixée → Donc succession d'opérations à respecter pour exécuter une opération, réaliser une expérience...

Technique(s) d'évaluation

Nom – noyau 2 : *technique(s) d'évaluation* → valeur fixée → Donc procédé professionnel, savoir-faire... Donc habileté, maîtrise...

Travail d'évaluation

Nom – noyau 1 : *travail d'évaluation* → fixer approximativement la valeur → Donc activité exercée pour parvenir à un résultat déterminé...

Nom – noyau 2 : *travail d'évaluation* → valeur fixée → Donc activité exercée pour parvenir à un résultat déterminé...

Pour les collocations *évaluation de la qualité*, *critère(s) d'évaluation* et *méthodologie(s) d'évaluation*, l'application du TSA montre que l'adjonction du collocatif à la base forme avec les stéréotypes proposés ou envisageables de *évaluer* et/ou de *évaluation(s)* un bloc d'argumentation interne (H31.1). Par conséquent, nous concluons pour chacune de ces collocations à la concrétisation d'un PA dans le discours (H31.2). Il s'agit de collocations argumentatives formant un bloc d'argumentation externe. Concernant les toutes les autres collocations, l'application du TSA montre que l'adjonction du collocatif à la base forme un déploiement argumentatif possible (H32.1). Par conséquent, nous concluons pour chacune de ces collocations à un mécanisme de motivation argumentative (H32.2).

Nous appliquons le TEA pour statuer sur leur axiologisation intrinsèque :

Évaluer la qualité → Donc améliorer l'analyse

Évaluer la qualité → *Pourtant améliorer l'analyse

Évaluer la qualité → Donc altérer l'analyse

Évaluer la qualité → *Pourtant altérer l'analyse

Évaluation de la qualité → Donc amélioration de l'analyse

Évaluation de la qualité → *Pourtant amélioration de l'analyse

Évaluation de la qualité → Donc altération de l'analyse

Évaluation de la qualité → *Pourtant altération de l'analyse

Critère(s) d'évaluation → Donc amélioration de l'analyse

Critère(s) d'évaluation → *Pourtant amélioration de l'analyse

Critère(s) d'évaluation → Donc altération de l'analyse

Critère(s) d'évaluation → *Pourtant altération de l'analyse

Processus d'évaluation → Donc amélioration de l'analyse

Processus d'évaluation → *Pourtant amélioration de l'analyse

Processus d'évaluation → Donc altération de l'analyse

Processus d'évaluation → *Pourtant altération de l'analyse

Dans cet ensemble de collocations, soit la base occupe la position tête, soit c'est le collocatif. Tous les blocs d'argumentations déployés sont normatifs, quelle que soit la SL mobilisée. Toutefois, ces alternances n'ont aucun effet sur l'axiologisation des collocations, car l'application du TEA montre que tous les enchaînements argumentatifs sont sémantiquement acceptables, ce qui signifie que chacun de ces PA et déploiements argumentatifs conserve son potentiel axiologique bivalent, en l'absence de tout élément co- ou contextuel.

Au niveau des mécanismes macro-sémantiques d'axiologisation de la signification (H41), la juxtaposition au sein d'une même collocation de deux unités lexicales bivalentes génère une collocation bivalente. Or aucun des mécanismes macro-sémantiques dont nous disposons ne permet de rendre compte de ce phénomène. Par défaut, nous pourrions statuer sur un mécanisme de neutralisation des valeurs axiologiques comme pour les collocatifs monovalents positifs. Mais dans les cas présents ni la base ni le collocatif ne perd sa « coloration » axiologique. Aussi, nous introduisons un nouveau mécanisme macro-sémantique dit de stabilisation des valeurs axiologiques bivalentes portées par *évaluer* et *évaluation(s)* pour expliciter avec justesse ce type de relation entre la base et le collocatif⁸. Nous attribuons ce mécanisme à toutes les collocations ci-dessus analysées.

⁸ Ce mécanisme macro-sémantique n'est pas réellement transposable au niveau discursif, dans la mesure où l'information co- et contextuelle reste rarement bivalente (*résultat de l'évaluation de la*

Au niveau discursif (H42), l'observation co- et contextuelle des collocations fait apparaître l'existence de deux mécanismes, l'un de renforcement négatif (avec toujours difficile), l'autre d'activation le plus souvent positive (avec *permettre, obtenir, intéressant, facteur important, données abondantes, grande qualité, réaliser...*) et occasionnellement négative (avec *pas possible* et l'association de *rester* et *important*).

Renforcement négatif

Ex : *il est toujours difficile d'évaluer la qualité d'un étiquetage* (AC_0056)

Activation positive

Ex : *la perplexité est un des critères qui permettent d'évaluer la qualité d'un modèle syntaxique* (AC_0011)

Ex : *une évaluation des performances obtenues par ce système et une comparaison aux autres approches de l'extraction sont réalisées en seconde partie* (AC_0083)

Ex : *il serait intéressant d'en évaluer les résultats sur des langues non apparentées, pour des textes de spécialité* (AR_0062)

Ex : *l'auto-évaluation du modèle est un facteur important* (AC_0050)

Ex : *le projet DISC a constitué de manière systématique et ascendante un ensemble de critères d'évaluation* (AR_0054)

Ex : *nous avons pour cela proposé un cadre applicatif ainsi qu'une méthodologie d'évaluation d'un tel système* (AC_0043)

Ex : *lorsque le paradigme d'évaluation utilise une méthodologie d'évaluation quantitative boîte noire (Sparck-Jones & Galliers, 1995), les données produites sont en général abondantes* (AC_0117)

Ex : *la campagne GRACE et l'expérience MULTITAG ont prouvé que le paradigme d'évaluation [...] peut aussi servir à produire de manière économique des ressources linguistiques validées de grande qualité* (AC_0117)

Ex : *une autre évolution marquante est l'adaptation croissante à d'autres tâches de TAL [...] des procédures d'évaluation développées initialement en recherche d'information* (AR_0075)

Ex : *l'évaluation du processus d'extraction a été réalisée à partir d'un corpus* (AR_0015)

Ex : *par exemple, il est indéniable que les grands programmes d'évaluation - tels que ceux de la (D)ARPA américaine (Pallett et al., 1994) par exemple [...] ont eu une influence très positive sur les recherches en CHM orale* (AR_0002)

Ex : *à ce stade de l'étude, nous sommes en mesure de lancer le processus d'évaluation* (AR_0065)

Ex : *MADCOW a été le premier à proposer un protocole d'évaluation de la compréhension* (AR_0054)

Ex : *le test et la validation des techniques d'évaluation relative sont en cours et donnent des résultats préliminaires encourageants* (AC_0023)

Activation négative

Ex : *il n'est donc pas possible de faire un programme d'évaluation qui prenne en compte tous les critères d'un expert pour décider de la qualité d'un index* (AR_0013)

Ex : *bien sûr, il reste encore un important travail d'évaluation comparative* (AC_0073)

En présence d'un potentiel axiologique bivalent, l'activation positive ou négative se fait donc par relais co- ou contextuel.

qualité), mais plutôt positive (*bonne évaluation de la qualité*), négative (*mauvaise évaluation de la qualité*) ou neutre (*décrire un processus d'évaluation*).

3.1.3.1.4 Les neutres

L'application du TSA concerne les collocations *évaluation des algorithmes*, *évaluation des méthodes*, *évaluation des systèmes*, *mesure(s) d'évaluation*, *outil(s) d'évaluation* et *phase(s) d'évaluation*.

Évaluation des algorithmes

Nom/verbe – noyau 1 : *évaluation des algorithmes* → fixer approximativement la valeur → Donc calcul, enchaînement des actions nécessaires à l'accomplissement d'une tâche...

Nom – noyau 2 : *évaluation des algorithmes* → valeur fixée → Donc calcul, enchaînement des actions nécessaires à l'accomplissement d'une tâche...

Verbe – noyau 2 : *évaluer les algorithmes* → déterminer par le calcul → Donc calcul, enchaînement des actions nécessaires à l'accomplissement d'une tâche...

Évaluation des méthodes

Nom/verbe – noyau 1 : *évaluation des méthodes* → fixer approximativement la valeur → Donc conduite et expression de la pensée... Donc procédé d'investigation, moyen...

Nom – noyau 2 : *évaluation des méthodes* → valeur fixée → Donc conduite et expression de la pensée... Donc procédé d'investigation, moyen...

Verbe – noyau 2 : *évaluer les méthodes* → déterminer par le calcul → Donc conduite et expression de la pensée... Donc procédé d'investigation, moyen...

Évaluation des systèmes

Nom/verbe – noyau 1 : *évaluation des systèmes* → fixer approximativement la valeur → Donc traitements matériels et logiciels de données...

Nom – noyau 2 : *évaluation des systèmes* → valeur fixée → Donc traitements matériels et logiciels de données...

Verbe – noyau 2 : *évaluer les systèmes* → déterminer par le calcul → Donc traitements matériels et logiciels de données...

Mesure(s) d'évaluation

Nom – noyau 1 : *évaluation des systèmes* → fixer approximativement la valeur → Donc évaluation, comparaison de la grandeur, de la quantité...

Nom – noyau 2 : *mesure(s) d'évaluation* → valeur fixée → Donc évaluation, comparaison de la grandeur, de la quantité...

Outil(s) d'évaluation

Nom – noyau 1 : *outils d'évaluation* → fixer approximativement la valeur → Donc objet fabriqué pour exécuter un travail...

Nom – noyau 2 : *outils d'évaluation* → valeur fixée → Donc objet fabriqué pour exécuter un travail...

Phase(s) d'évaluation

Contrairement à la majorité des collocations lexicales recueillies, dans *phase(s) d'évaluation*, *évaluation* est employé uniquement au sens du noyau 1.

Nom – noyau 1 : *phase(s) d'évaluation* → fixer approximativement la valeur → Donc états successifs... Donc stade, étape...

Pour les collocations *évaluation des algorithmes* et *mesure(s) d'évaluation*, l'application du TSA montre que l'adjonction du collocatif à la base forme avec les stéréotypes proposés ou envisageables de *évaluer* et/ou de *évaluation(s)* un bloc d'argumentation interne (H31.1). Par conséquent, nous concluons pour chacune de ces collocations à la concrétisation d'un PA dans le discours (H31.2). Il s'agit de collocations argumentatives formant un bloc d'argumentation externe. Pour les autres collocations, l'application du TSA montre que l'adjonction du collocatif à la base forme un déploiement argumentatif possible (H32.1). Par conséquent, nous concluons pour chacune de ces collocations à un mécanisme de motivation argumentative (H32.2).

Nous appliquons le TEA pour statuer sur leur axiologisation intrinsèque :

Évaluer les algorithmes → Donc améliorer l'analyse

Évaluer les algorithmes → *Pourtant améliorer l'analyse

Évaluer les algorithmes → Donc altérer l'analyse

Évaluer les algorithmes → *Pourtant altérer l'analyse

Évaluation des algorithmes → Donc amélioration de l'analyse

Évaluation des algorithmes → *Pourtant amélioration de l'analyse

Évaluation des algorithmes → Donc altération de l'analyse

Évaluation des algorithmes → *Pourtant altération de l'analyse

Évaluation des méthodes → Donc amélioration de l'analyse

Évaluation des méthodes → *Pourtant amélioration de l'analyse

Évaluation des méthodes → Donc altération de l'analyse

Évaluation des méthodes → *Pourtant altération de l'analyse

Dans cet ensemble de collocations, soit la base occupe la position tête, soit c'est le collocatif. Tous les blocs d'argumentations déployés sont normatifs, quelle que soit la SL mobilisée. Toutefois, comme pour les collocatifs bivalents, ces alternances n'ont aucun effet sur l'axiologisation des collocations, car l'application du TEA montre que tous les enchaînements argumentatifs déployés sont sémantiquement acceptables, ce qui signifie que chacun de ces déploiements argumentatifs est axiologiquement bivalent, en l'absence de tout élément co- ou contextuel.

Au niveau des mécanismes macro-sémantiques d'axiologisation de la signification (H41), la présence au sein d'une même collocation d'une unité lexicale bivalente et d'une unité neutre génère une collocation bivalente. Nous statuons donc sur un mécanisme de neutralisation des potentiels axiologiques bivalents. Nous attribuons ce mécanisme à toutes les collocations ci-dessus analysées.

Au niveau discursif, l'étude co- et contextuelle des collocations fait apparaître l'existence de deux mécanismes, l'un de renforcement systématiquement positif (avec *toujours important*) et l'autre d'activation également positive (avec *important d'examiner*, *s'accorder sur*, *utiliser*, *être disponible* et *permettre*, *mettre en évidence*), occasionnellement négative (avec *problèmes mis en évidence*).

Renforcement positif

Ex : *il est toujours important d'avoir des descriptions de méthodes d'évaluation appliquées à des systèmes* (AR_0013)

Activation positive

Ex : nous avons appliqué les méthodes de segmentation de *SEGCORLEX* et *SEGAPSITH* sur une tâche classique d'évaluation des algorithmes de segmentation thématique (AR_0005)

Ex : l'évaluation des méthodes de filtrage syntaxique des cooccurrences a été effectuée pour la tâche de la recherche documentaire (AC_0049)

Ex : il est donc important d'examiner les différents aspects de l'évaluation des systèmes de QR (AR_0074)

Ex : on s'accorde en effet sur les mesures d'évaluation des systèmes de recherche d'information, principalement la précision et le rappel (AR_0074)

Ex : pour évaluer un critère sur le corpus, nous utilisons une méthode d'évaluation croisée k fois (*k-fold cross validation* en anglais), conformément à l'usage commun, $k=10$ dans notre expérience (AC_0079)

Ex : les outils d'évaluation seront disponibles au printemps prochain (projet européen *ELSE*), permettant entre autres d'évaluer une portion de texte quelconque par rapport à la référence (AC_0016)

Activation négative

Ex : cependant, un certain nombre de problèmes sont mis en évidence dans la phase d'évaluation (AR_0058)

3.1.3.2 Le collocatif est un adjectif

3.1.3.2.1 Les monovalents positifs

Le TSA est appliqué aux collocations *évaluation fine*, *évaluation objective*, *évaluation qualitative* et *intéressant d'évaluer*.

Évaluation fine

Nom/verbe – noyau 1 : *évaluation fine* → fixer approximativement la valeur → *Donc précision, exactitude...

Nom/verbe – noyau 1 : *évaluation fine* → fixer approximativement la valeur → Pourtant précision, exactitude...

Nom – noyau 2 : *évaluation fine* → valeur fixée → Donc précision, exactitude...

Évaluation objective

Nom/verbe – noyau 1 : *évaluation objective* → fixer approximativement la valeur → *Donc réel... *Donc indépendant de toute préférence subjective...

Nom/verbe – noyau 1 : *évaluation objective* → fixer approximativement la valeur → Pourtant réel... Pourtant indépendant de toute préférence subjective...

Nom – noyau 2 : *évaluation objective* → valeur fixée → Donc réel... Donc indépendant de toute préférence subjective...

Évaluation qualitative

Dans la langue générale, *qualitatif-ive* est relatif à *qualité* et peut donc être comme le nom soit bivalent (employé au sens de *caractéristique*, *valeur*), soit monovalent positif (*bonne qualité*, *qualité positive*). Les occurrences du mot dans le corpus TAL(N) font apparaître un

emploi positif systématique. C'est pourquoi, de même que pour qualité, nous choisissons d'analyser *qualitative* dans cette section et de ne faire figurer que le stéréotype positif.

Nom – noyau 1 : *évaluation qualitative* → fixer approximativement la valeur → *Donc bonne caractéristique, valeur positive...

Nom – noyau 1 : *évaluation qualitative* → fixer approximativement la valeur → Pourtant bonne caractéristique, valeur positive...

Nom – noyau 2 : *évaluation qualitative* → valeur fixée → Donc bonne caractéristique, valeur positive...

Intéressant d'évaluer

Verbe – noyau 1 : *intéressant d'évaluer* → fixer approximativement la valeur → Donc captivant... Donc important... Donc avantageux, valable...

Verbe – noyau 2 : *intéressant d'évaluer* → déterminer par le calcul → Donc captivant... Donc important... Donc avantageux, valable...

L'application du TSA montre que l'adjonction du collocatif à la base forme avec les stéréotypes proposés ou envisageables de *évaluer* et/ou de *évaluation(s)* un bloc d'argumentation interne (H31.1), déployé de façon transgressive avec *fine* et *objective*. Par conséquent, nous concluons pour chacune de ces collocations à la concrétisation d'un PA dans le discours (H31.2). Il s'agit de collocations argumentatives formant un bloc d'argumentation externe. Pour les collocations *évaluation qualitative* et *intéressant d'évaluer*, l'application du TSA montre que l'adjonction du collocatif à la base forme un déploiement argumentatif possible (H32.1). Il s'agit d'un mécanisme de motivation argumentative (H32.2).

Pour statuer sur l'axiologisation intrinsèque, nous appliquons le TEA :

Évaluer finement → Donc améliorer l'analyse

Évaluer finement → *Pourtant améliorer l'analyse

Évaluer finement → *Donc altérer l'analyse

Évaluer finement → Pourtant altérer l'analyse

Évaluation fine → Donc amélioration de l'analyse

Évaluation fine → *Pourtant amélioration de l'analyse

Évaluation fine → *Donc altération de l'analyse

Évaluation fine → Pourtant altération de l'analyse

Évaluation qualitative → Donc amélioration de l'analyse

Évaluation qualitative → *Pourtant amélioration de l'analyse

Évaluation qualitative → *Donc altération de l'analyse

Évaluation qualitative → Pourtant altération de l'analyse

Que la base soit en position tête ou pas, que le bloc d'argumentation déployé soit transgressif, ou qu'il corresponde seulement à un déploiement argumentatif possible, seuls les enchaînements argumentatifs normatif avec « améliorer – amélioration » et transgressif avec « altérer – altération » sont sémantiquement acceptables, ce qui signifie que les collocations sont intrinsèquement axiologisées positivement.

Au niveau des mécanismes macro-sémantiques d'axiologisation de la signification (H41), nous statuons sur un mécanisme d'activation positive des valeurs axiologiques bivalentes portées par *évaluer* et *évaluation(s)*. Nous attribuons ce mécanisme à toutes les collocations ci-dessus.

Au niveau discursif, l'étude co- et contextuelle des collocations fait apparaître l'existence de trois mécanismes : le renforcement (avec *primordial*), l'activation (avec *présenter*, *être aussi important*, *paraître*) et l'affaiblissement (avec *empêcher* et *se révéler délicat*).

Renforcement

Ex : les *évaluations objectives* des performances des systèmes sont primordiales pour diagnostiquer les capacités des différents modules (AR_0054)

Activation

Ex : l'article présente une *évaluation fine* de ces deux techniques (AC_0068)

Ex : mais l'*évaluation qualitative* est pour le moins aussi importante (AC_0013)

Ex : il nous a paru *intéressant d'évaluer* le nombre de candidats où l'adjectif relationnel n'est plus en concurrence avec un groupe prépositionnel effectivement présent dans le thesaurus (AC_0006)

Affaiblissement

Ex : ceci nous empêche, pour l'instant, d'*évaluer plus finement* les résultats, voire de nouveaux étiqueteurs (AC_0016)

Ex : pour cette raison, une *évaluation objective* et expérimentale se révèle assez délicate (AR_0045)

3.1.3.2.2 Le monovalent négatif

L'application du TSA concerne la collocation *difficile à évaluer*.

Difficile à évaluer

Verbe – noyau 1 : *difficile à évaluer* → fixer approximativement la valeur → Donc effort, peine... Donc laborieux, pénible...

Verbe – noyau 2 : *difficile à évaluer* → déterminer par le calcul → Donc effort, peine... Donc laborieux, pénible...

L'application du TSA montre que l'adjonction du collocatif *difficile* à la base *évaluer* forme un déploiement argumentatif possible (H32.1). Nous concluons pour *difficile à évaluer* à un mécanisme de motivation argumentative (H32.2).

Nous utilisons le TEA pour statuer sur son axiologisation intrinsèque (H32.3) :

Difficile à évaluer → Donc temps de travail élevé

Difficile à évaluer → *Pourtant temps de travail élevé

Difficile à évaluer → *Donc temps de travail réduit

Difficile à évaluer → Pourtant temps de travail réduit

Dans ce cas, le collocatif est en position tête et le bloc d'argumentation est normatif. Seuls les enchaînements normatif avec « travail élevé » et transgressif avec « travail réduit » sont sémantiquement acceptables, ce qui signifie que le déploiement argumentatif est intrinsèquement axiologisé négativement.

Au niveau des mécanismes macro-sémantiques d'axiologisation de la signification (H41), la polarité négative du collocatif contamine donc la bivalence de la base, ce qui nous permet de statuer sur un mécanisme d'activation des valeurs axiologiques bivalentes portées par *évaluer*.

Au niveau discursif, l'observation cotextuelle de la collocation fait apparaître un seul mécanisme de renforcement, avec *fastidieux*.

Renforcement

Ex : *l'énumération des concepts activés est vite fastidieuse et surtout **difficile à évaluer*** (AC_0042)

3.1.3.2.3 Les bivalents

L'application du TSA concerne les collocations *évaluation comparative*, *évaluation empirique*, *évaluation globale*, *évaluation manuelle* et *évaluations expérimentales*.

Évaluation comparative

Nom – noyau 1 : *évaluation comparative* → fixer approximativement la valeur → Donc établir une analyse, un jugement, un rapprochement...

Nom – noyau 2 : *évaluation empirique* → valeur fixée → Donc établir une analyse, un jugement, un rapprochement...

Évaluation empirique

Nom – noyau 1 : *évaluation empirique* → fixer approximativement la valeur → Donc qui résulte de l'expérience... Donc expérimental, pragmatique, pratique...

Nom – noyau 2 : *évaluation empirique* → valeur fixée → Donc qui résulte de l'expérience... Donc expérimental, pragmatique, pratique...

Évaluation globale

Nom – noyau 1 : *évaluation globale* → fixer approximativement la valeur → Donc considéré dans son ensemble... Donc complet, intégral...

Nom – noyau 2 : *évaluation globale* → valeur fixée → Donc considéré dans son ensemble... Donc complet, intégral...

Évaluation manuelle

Nom/verbe – noyau 1 : *évaluation manuelle* → fixer approximativement la valeur → *Donc en se servant de la main, en travaillant physiquement...

Nom/verbe – noyau 1 : *évaluation manuelle* → fixer approximativement la valeur → *Pourtant en se servant de la main, en travaillant physiquement...

Nom – noyau 2 : *évaluation manuelle* → valeur fixée → *Donc en se servant de la main, en travaillant physiquement...

Nom – noyau 2 : *évaluation manuelle* → valeur fixée → *Pourtant en se servant de la main, en travaillant physiquement...

Verbe – noyau 2 : *évaluer manuellement* → déterminer par le calcul → *Donc en se servant de la main, en travaillant physiquement...

Verbe – noyau 2 : *évaluer manuellement* → déterminer par le calcul → *Pourtant en se servant de la main, en travaillant physiquement...

Évaluations expérimentales

Nom/verbe – noyau 1 : *évaluations expérimentales* → fixer approximativement la valeur → Donc étude d'un phénomène provoqué pour observer les résultats...

Nom – noyau 2 : *évaluations expérimentales* → valeur fixée → Donc étude d'un phénomène provoqué pour observer les résultats...

Verbe – noyau 2 : *évaluer expérimentalement* → déterminer par le calcul → Donc de manière à étudier un phénomène provoqué pour observer les résultats...

L'application du TSA montre que l'adjonction du collocatif *comparative* à la base forme avec les stéréotypes proposés de *évaluer* et/ou de *évaluation(s)* un bloc d'argumentation interne (H31.1). Par conséquent, nous concluons à la concrétisation d'un PA dans le discours (H31.2). Cette collocation est argumentative. Pour les collocations *évaluation empirique*, *évaluation globale* et *évaluations expérimentales*, l'application du TSA montre que l'adjonction du collocatif à la base forme un déploiement argumentatif possible (H32.1). Par conséquent, nous concluons à un mécanisme de motivation argumentative (H32.2). Concernant la collocation *évaluation manuelle*, l'application du TSA montre que l'adjonction du collocatif à la base ne forme ni bloc d'argumentation interne, ni déploiement argumentatif possible (H33.1). Par conséquent, cette collocation est arbitraire (H33.2).

Nous appliquons le TEA pour statuer sur leur axiologisation intrinsèque :

Évaluation comparative → Donc amélioration de l'analyse

Évaluation comparative → *Pourtant amélioration de l'analyse

Évaluation comparative → Donc altération de l'analyse

Évaluation comparative → *Pourtant altération de l'analyse

Évaluation globale → Donc amélioration de l'analyse

Évaluation globale → *Pourtant amélioration de l'analyse

Évaluation globale → Donc altération de l'analyse

Évaluation globale → *Pourtant altération de l'analyse

Évaluer manuellement → Donc améliorer l'analyse

Évaluer manuellement → *Pourtant améliorer l'analyse

Évaluer manuellement → Donc altérer l'analyse

Évaluer manuellement → *Pourtant altérer l'analyse

Évaluation manuelle → Donc amélioration de l'analyse

Évaluation manuelle → *Pourtant amélioration de l'analyse

Évaluation manuelle → Donc altération de l'analyse

Évaluation manuelle → *Pourtant altération de l'analyse

Dans cet ensemble de collocations, la base occupe systématiquement la position tête. L'application du TEA montre que les enchaînements argumentatifs normatifs sont sémantiquement acceptables, ce qui signifie que chacun de ces déploiements conserve son potentiel axiologique bivalent, en l'absence de tout élément co- ou contextuel. Il en va de même pour la collocation arbitraire.

Au niveau des mécanismes macro-sémantiques d'axiologisation de la signification (H41), la juxtaposition au sein d'une même collocation de deux unités lexicales bivalentes génère une collocation bivalente. Comme pour les collocations construites avec un nom, nous statuons sur un mécanisme macro-sémantique dit de neutralisation-stabilisation des valeurs axiologiques bivalentes portées par *évaluer* et *évaluation(s)*.

Au niveau discursif (H42), l'observation co- et contextuelle des collocations fait apparaître l'existence de deux mécanismes, l'un d'activation (voire de renforcement avec la présence de *évidemment*) positive ou négative (avec *intéressant*, *expérience de validation* ou *important*

travail, coûteux et se heurter), l'autre de neutralisation (avec *consacrer, discuter et présenter*).

Activation positive

Ex : *il serait maintenant évidemment intéressant de mener des **évaluations comparatives** avec d'autres systèmes sur la même tâche* (AR_0054)

Ex : *l'**évaluation globale** du nombre d'erreurs reste le facteur orientant de manière prédominante les développements d'applications de dictée vocale* (AR_0034)

Ex : *la dernière expérience de validation de FLEMM a été une **évaluation manuelle** effectuée à l'IRIN (Nantes) à partir d'un corpus en français* (AR_0017)

Activation négative

Ex : *bien sûr, il reste encore un important travail d'**évaluation comparative*** (AC_0073)

Ex : *la mise en place des ensembles d'unités de références est certes semi-automatique mais reste moins coûteuse que les **évaluations manuelles*** (AR_0054)

Ex : *l'**évaluation globale** d'un système tel que TOPICOLL se heurte à un problème* (AC_0060)

Neutralisation

Ex : *une sous-partie est consacrée à l'**évaluation empirique** de la complexité algorithmique* (AC_0038)

Ex : *nous discutons de l'**évaluation globale** des familles morphologiques trouvées* (AC_0010)

Ex : *enfin, deux **évaluations expérimentales** sont présentées en section 5* (AR_0018)

La possibilité d'un mécanisme discursif de neutralisation des collocations bivalentes va à l'encontre de notre hypothèse H42, puisque nous postulons qu'en l'absence d'orientation axiologique intrinsèque à la collocation la sélection de la polarité positive ou négative se fait par relais co- ou contextuel. Selon cette hypothèse, nous nous attendons donc à ce que la bivalence des collocations ci-dessus analysées soit rompue et instanciée positivement ou négativement par les éléments cotextuels, ce qui, à l'évidence, n'est pas le cas ici. Toutefois, ces mécanismes discursifs de neutralisation cotextuelle correspondent à des énoncés d'un genre particulier, situés en introduction des articles et visant à en présenter la structure. Aussi, cette neutralité est ensuite rompue, lorsqu'il s'agit de décrire le type d'évaluation retenue et d'en exposer soit les mérites, soit les méfaits.

3.1.3.2.4 Les neutres

Le TSA est appliqué aux collocations *évaluation automatique* et *évaluation quantitative*.

Évaluation automatique

Nom/verbe – noyau 1 : *évaluation automatique* → fixer approximativement la valeur → *Donc fonctionne de lui-même... *Donc de manière mécanique, informatisée, programmée...

Nom/verbe – noyau 1 : *évaluation automatique* → fixer approximativement la valeur → *Pourtant fonctionne de lui-même... *Pourtant de manière mécanique, informatisée, programmée...

Nom – noyau 2 : *évaluation automatique* → valeur fixée → *Donc fonctionne de lui-même... *Donc de manière mécanique, informatisée, programmée...

Nom – noyau 2 : *évaluation automatique* → valeur fixée → *Pourtant fonctionne de lui-même... *Pourtant de manière mécanique, informatisée, programmée...

Verbe – noyau 2 : *évaluer de manière automatique* → déterminer par le calcul → *Donc fonctionne de lui-même... *Donc de manière mécanique, informatisée, programmée...

Verbe – noyau 2 : *évaluer de manière automatique* → déterminer par le calcul → *Pourtant fonctionne de lui-même... *Pourtant de manière mécanique, informatisée, programmée...

Évaluation quantitative

Nom – noyau 2 : *évaluation quantitative* → valeur fixée → Donc (concerne) un nombre, une mesure, un poids...

L'application du TSA montre que l'adjonction du collocatif *automatique* à la base *évaluer* et/ou de *évaluation(s)* ne forme ni bloc d'argumentation, ni déploiement argumentatif possible (H33.1). Par conséquent, nous concluons à une collocation arbitraire (H33.2). Concernant la collocation *évaluation quantitative*, l'application du TSA montre que l'adjonction du collocatif à la base forme avec les stéréotypes proposés ou envisageables de *évaluation* un bloc d'argumentation interne (H31.1). Par conséquent, nous concluons à la concrétisation d'un PA dans le discours (H31.2). Il s'agit d'une collocation argumentative.

Nous appliquons le TEA pour statuer sur leur axiologisation intrinsèque :

Évaluer de manière automatique → Donc améliorer l'analyse

Évaluer de manière automatique → *Pourtant améliorer l'analyse

Évaluer de manière automatique → Donc altérer l'analyse

Évaluer de manière automatique → *Pourtant altérer l'analyse

Évaluation automatique → Donc amélioration de l'analyse

Évaluation automatique → *Pourtant amélioration de l'analyse

Évaluation automatique → Donc altération de l'analyse

Évaluation automatique → *Pourtant altération de l'analyse

Évaluation quantitative → Donc amélioration de l'analyse

Évaluation quantitative → *Pourtant amélioration de l'analyse

Évaluation quantitative → Donc altération de l'analyse

Évaluation quantitative → *Pourtant altération de l'analyse

Dans ces deux cas, la base occupe la position tête de la collocation et les blocs d'argumentations déployés sont normatifs, quelle que soit la SL mobilisée. L'application du TEA montre que tous les enchaînements en « donc » sont acceptables, ce qui signifie que ces deux PA sont axiologiquement bivalents.

Au niveau des mécanismes macro-sémantiques d'axiologisation de la signification (H41), la présence au sein d'une même collocation d'une unité lexicale bivalente et d'une unité neutre génère une collocation bivalente. Nous statuons donc sur un mécanisme de neutralisation des potentiels axiologiques bivalents.

Au niveau discursif, l'observation co- et contextuelle des collocations fait apparaître l'existence d'un seul mécanisme d'activation soit positive (avec *évaluation par des experts* et *pertinence des résultats obtenus*), soit négative (avec l'association de *ne pas toujours être souhaitable* et *trop complexe*).

Activation positive

Ex : *une évaluation automatique* peut être équivalente à *une évaluation par des experts* (AR_0013)

Ex : nous avons réalisé une *évaluation quantitative* de la pertinence des résultats obtenus par la combinaison (AC_0096)

Activation négative

Ex : une *évaluation automatique* peut aussi ne pas toujours être souhaitable lorsque la décision est trop complexe (AR_0013)

3.2 Résultat(s)

Le corpus TAL(N) comporte 306 occurrences de *résultat* au singulier et 1291 occurrences du mot au pluriel.

3.2.1 Dépouillement des données

3.2.1.1 Nettoyage

Selon le critère statistique, nous éliminons les candidats trop peu fréquents. Nous retirons donc tous les apax, les candidats de fréquence deux et les cooccurents n'apparaissant pas dans trois articles distincts.

Ex : *résultat chiffré* (1), *résultats comparables* (2), *obtention d'un résultat* (1), *production de résultats* (1), *résultats spectaculaires* (1).

Selon le critère de pertinence, nous laissons de côté toutes les collocations grammaticales et autres constructions dont le collocatif ne présente pas de contenu sémantique.

Ex : *résultat(s) de/du/des*.

Selon le critère lexical, une seule lexie construite sur *résultat(s)* et rassemblant trois occurrences a été trouvée en corpus (AR_0012, AR_0068). Il s'agit de *résultat brut*, restreinte au domaine scientifique et définie dans le GR par résultat obtenu sans les corrections nécessaires. Fait brut, résultat brut, non analysé.

De même, selon le critère lexicographique, le corpus TAL(N) comporte au moins deux termes complexes *automate résultat* (5), *arbre résultat* (2). Toutefois, ces termes ne sont pas mentionnés ni dans le GDT ni dans Eurodicautom, ils ont été validés par des experts du domaine.

3.2.1.2 Liste des collocations lexicales

Cette sous-section présente les quatre tableaux récapitulatifs des collocations lexicales construites sur *résultat(s)*. Les deux premiers tableaux correspondent au contexte droit du substantif et les deux derniers, à son contexte gauche.

COLLOCATIONS LEXICALES	FREQ.	AUTRES FORMES	FREQ.	TOTAL
Résultat brut	1	-Résultats Bruts	2	3
Résultat complet	2	-Résultats complets	3	5
Résultat correct	2	-Résultats corrects	5	7
Résultat d'analyse	1	-R. de l'analyse -R. d'une analyse -Résultats de l'analyse -Résultats [...] analyses	7 2 16 8	37
Résultat d'une application	1	-R. de l'application	4	5
Résultat d'une opération	1	-R. [...] opération	3	4
Résultat décevant	1	-Résultats décevants	2	3
Résultat de l'approche	1	-Résultats de l'approche	2	3
Résultat de la méthode	2	-Résultats de la méthode -Résultats [...] méthode(s)	4 6	12
Résultat de la recherche	2	-Résultats de la recherche -Résultats [...] recherches	5 2	9
Résultat du processus	1	-R. [...] processus	2	3
Résultat empirique	1	-Résultats empiriques	3	4
Résultat encourageant	1	-R. [...] encourageant -Résultats [...] encourageants	3 4	8
Résultat final	10	-Résultats finaux	4	14
Résultat global	2	-Résultats globaux	11	13
Résultat intéressant	2	-R. [...] intéressant -Résultats intéressants -Résultats [...] intéressants	3 6 5	16
Résultat positif	2	-Résultats [...] positifs	3	5
Résultat satisfaisant	2	-Résultats satisfaisants	11	13
Résultat théorique	1	-Résultats théoriques	4	5

Tableau 28 : Collocations lexicales contexte droit de résultat(s)

COLLOCATIONS LEXICALES	FREQ.	AUTRES FORMES	FREQ.	TOTAL
Résultats bons	-	-R. [...] bons -Bon résultat -Bons R.	5 2 36	43
Résultats de cette étude	5	-	-	5
Résultats des expériences	2	-R. [...] expériences	13	15
Résultats des travaux	3	-	-	3
Résultats du système	4	-R. [...] système	8	12
Résultats du test	3	-R. [...] test	3	6
Résultats expérimentaux	6	-	-	6
Résultats mauvais	-	-R. [...] mauvais -Mauvais R.	1 2	3
Résultats meilleurs	1	-R [...] meilleurs -Meilleur résultat -Meilleurs R.	9 5 45	60
Résultats numériques	6	-	-	6
Résultats pertinents	4	-Pertinence du R. -Pertinence des R. -Pertinence [...] R.	1 3 1	9
Résultats préliminaires	7	-	-	7
Résultats prometteurs	4	-R. [...] prometteurs	2	6
Résultats quantitatifs	6	-	-	6
Résultats significatifs	2	-R. [...] significatifs	1	3
Résultats statistiques	3	-	-	3
Résultats supérieurs	2	-R. [...] supérieurs	3	5

Tableau 29 : Collocations lexicales contexte droit de résultats

COLLOCATIONS LEXICALES	FREQ.	AUTRES FORMES	FREQ.	TOTAL
Analyse du résultat	1	-Analyse des résultats -Analyse [...] des R.	6 5	12
Cohérence du résultat	2	-Cohérence des résultats	2	4

Tableau 30 : Collocations lexicales contexte gauche de résultat(s)

COLLOCATIONS LEXICALES	FREQ.	AUTRES FORMES	FREQ.	TOTAL
Comparaison des résultats	4	-Comparer + R.	21	25
Dégradation des résultats	2	-R. + dégrader	2	4
Excellents résultats	3	-	-	3
Exploitation des résultats	2	-Exploiter + R.	7	9
Fiabilité des résultats	4	-Fiabilité [...] des R.	1	5
Interprétation des résultats	5	-Interpréter + R.	4	9
Qualité des résultats	12	-Qualité [...] R. -Qualité du résultat	1 4	17
Validité des résultats	1	-Validité [...] des R.	4	5

Tableau 31 : Collocations lexicales contexte gauche de résultats

3.2.2 Signification lexicale de résultat

Dans le LL, *résultat* bénéficie d'une entrée à part entière englobant également le verbe transitif *résulter* et l'adjectif *résultant-e* (cf. Annexe 9). L'article se compose de trois définitions, deux pour le nom singulier, une pour le nom employé au pluriel relative au domaine comptable – Solde du compte des profits et pertes (*les résultats de l'exercice ont été bons*). Nous ignorons cette définition pour la construction de la SL.

Définition 1 : Ce qui arrive, se produit à la suite d'une action, d'un événement, de l'application d'un principe, d'une opération mathématique.

→ conclusion, dénouement, effet (*résultat d'une négociation*)

→ réussite, succès, conséquence, échec (*résultat inespéré ; résultats lamentables ; résultat d'une expérience*)

→ somme, reste, produit, quotient (*résultat d'une addition/soustraction ; résultat d'une multiplication/division*)

Définition 2 : Réussite ou échec à un concours, à un examen, à une compétition ; liste des candidats qui ont réussi (*commenter les résultats d'un match*).

Le GR propose une entrée pour *résultat* (cf. Annexe 9). L'article propose quatre catégories de définition : langue classique, langue moderne et deux autres d'ordre général. Nous laissons la première catégorie de côté, ainsi que toutes les constructions grammaticales lexicalisées et les lexies (locutions) associées à la langue moderne : être le résultat de qqc., sans résultat, en résultat.

Définition 1 : Ce qui arrive, se produit, commence à exister à la suite et comme effet de qqch. (activité, événement, phénomène, opération, principe).

Ce que produit une activité consciente dirigée vers une fin ; cette même fin.

→ conséquence, contrecoup, effet, événement, fin, issue, solution, succès, suite (*résultat d'un excès de travail, donner un résultat*)

→ aboutir, efficace, réussir, progrès, échec, méfait (*rechercher un résultat ; viser un résultat ; atteindre un résultat ; approcher du résultat ; résultats désastreux*)

→ instrument, moyen, opération (*obtenir un résultat ; produire un résultat*)

→ (souvent au pl.) fructifier, réussir, recueillir (*exiger des résultats ; résultats fructueux*)

Définition 2 : Solution (d'un problème). Phase ultime d'un calcul ; troisième élément associé à un couple dans une application, une opération arithmétique.

→ produit, quotient, reste, somme, réponse (*résultat d'une opération*)

Définition 3 : Admission ou non-admission, réussite ou échec (à un examen, un concours) ; liste de ceux qui ont réussi.

→ réussir, gagner, échouer (*affichage des résultats ; résultats des élections ; résultat définitifs*)

La définition 2 du GR étant présente dans le LL, mais contenue dans la première, nous choisissons également de la regrouper dans un seul noyau, mais de la mentionner explicitement dans les stéréotypes.

Les deux dictionnaires décrivent *résultat* comme un nom masculin. Cette unité lexicale appartient à la zone modale de l'axiologique, conformément aux synonymes proposés dont l'orientation est soit positive (*succès*), soit négative (*échec*), ce qui fait de lui un bivalent.

Figure ci-dessous le tableau de synthèse de la signification lexicale du substantif.

RESULTAT(S)			
TRAITS NECESSAIRES DE CATEGORISATION (TNC)			
TYPE GRAMMATICAL		ZONE MODALE	ORIENTATION AXIOLOGIQUE
Nom – Masculin		Axiologique	Bivalent
ENTITES SEMANTIQUES			ENTITES DISCURSIVES
NOYAU		STEREOTYPES	POSSIBLES ARGUMENTATIFS
1	Ce qui arrive, se produit à la suite et comme effet de qqc.	Donc conséquence, issue... Donc échec, méfait, réussite, succès... Donc aboutissement, efficacité... Donc instrument, moyen, opération... Donc produit, quotient, reste...	R. donc issue d'un travail R. donc succès inespéré R. donc viser un progrès R. donc obtention un score R. donc solution d'une opération
2	Réussite ou échec (à un concours, un examen) ; liste de ceux qui ont réussi	Donc aboutissement, réalisation... Donc chance, gain, succès... Donc défaite, faillite, infortune... Donc affichage, présentation, commentaire, discussion...	R. donc réalisation définitive R. donc succès électoral R. donc infortune R. donc affichage des lauréats

Tableau 32 : Signification lexicale de résultat(s)

3.2.3 Analyse des données

3.2.3.1 Le collocatif est un nom

3.2.3.1.1 Les monovalents positifs

Le TSA est appliqué aux collocations suivantes : *résultat de la recherche*, *résultats de cette étude*, *analyse du résultat*, *cohérence du résultat*, *exploitation des résultats*, *fiabilité des résultats*, *interprétation des résultats* et *validité des résultats*.

Résultat de la recherche

Malgré la présence de *effort* dans les éléments de SL de *recherche*, nous choisissons de classer cette collocation dans les monovalents positifs eu égard à l'objectif de l'effort orienté vers l'activité de *trouver*, indiscutablement positive.

SL – noyau 1 : *résultat de la recherche* → ce qui arrive à la suite et comme effet de qqc. → Donc effort pour trouver qqc. ...

Résultats de cette étude

SL – noyau 1 : *résultat de cette étude* → ce qui arrive à la suite et comme effet de qqc. → Donc application méthodique de l'esprit... Donc effort intellectuel orienté vers l'observation...

Analyse du résultat

SL – noyau 1 : *analyse du résultat* → ce qui arrive à la suite et comme effet de qqc. → Donc étude détaillée...

Cohérence du résultat

SL – noyau 1 : *cohérence du résultat* → ce qui arrive à la suite et comme effet de qqc. → Donc harmonieux, logique, moins contradictoire...

Exploitation des résultats

SL – noyau 1 : *exploitation des résultats* → ce qui arrive à la suite et comme effet de qqc. → Donc utilisation méthodique...

Fiabilité des résultats

SL – noyau 1 : *fiabilité des résultats* → ce qui arrive à la suite et comme effet de qqc. → Donc crédibilité, vraisemblance...

Interprétation des résultats

SL – noyau 1 : *interprétation des résultats* → ce qui arrive à la suite et comme effet de qqc. → *Donc explication, signification claire...

SL – noyau 1 : *interprétation des résultats* → ce qui arrive à la suite et comme effet de qqc. → *Pourtant explication, signification claire...

Validité des résultats

SL – noyau 1 : *validité des résultats* → ce qui arrive à la suite et comme effet de qqc. → Donc bien-fondé, valeur...

L'application du TSA montre que l'adjonction du collocatif à la base *résultat(s)* forme un déploiement argumentatif possible (H32.1). Par conséquent, nous concluons pour chacune de ces collocations à un mécanisme de motivation argumentative (H32.2). En revanche, concernant *interprétation des résultats*, l'application du TSA montre que l'adjonction du collocatif à la base *résultat(s)* ne forme ni un bloc d'argumentation, ni un déploiement argumentatif possible (H33.1). Par conséquent, nous concluons à une collocation arbitraire (H33.2).

Pour statuer sur leur axiologisation intrinsèque, nous recourons au TEA :

Résultat de la recherche → Donc amélioration de l'analyse

Résultat de la recherche → *Pourtant amélioration de l'analyse

Résultat de la recherche → Donc altération de l'analyse

Résultat de la recherche → *Pourtant altération de l'analyse

Cohérence du résultat → Donc amélioration de l'analyse

Cohérence du résultat → *Pourtant amélioration de l'analyse

Cohérence du résultat → *Donc altération de l'analyse

Cohérence du résultat → Pourtant altération de l'analyse

Interprétation des résultats → Donc amélioration de l'analyse

Interprétation des résultats → *Pourtant amélioration de l'analyse

Interprétation des résultats → Donc altération de l'analyse

Interprétation des résultats → *Pourtant altération de l'analyse

Dans cette série de données, tantôt la base occupe la position tête de la collocation, tantôt c'est le collocatif, et tous les déploiements argumentatifs sont normatifs. Nous observons une variation de l'axiologisation intrinsèque de la collocation en fonction de l'élément tête.

Lorsque la base est en position tête comme dans *résultat de la recherche*, seuls les blocs d'argumentations normatifs en DC (Donc) « amélioration » et « altération » sont

sémantiquement valides, ce qui signifie que le déploiement argumentatif est intrinsèquement bivalent. La polarité positive du collocatif ne se transmet pas à la base.

Lorsque le collocatif est en position tête comme dans *cohérence du résultat*, seuls les blocs d'argumentations normatif avec « amélioration » et transgressif avec « altération » sont sémantiquement acceptables, ce qui signifie que l'argumentation est intrinsèquement axiologisée positivement. La polarité positive du collocatif semble se transmettre à la base.

Concernant *interprétation des résultats*, seuls les enchaînements en « donc » sont acceptables, la collocation est donc bivalente.

Au niveau des mécanismes macro-sémantiques d'axiologisation de la signification (H41), dans le premier cas, la polarité positive du collocatif ne contamine pas la bivalence de la base, ce qui nous permet de statuer sur un mécanisme de neutralisation des valeurs axiologiques bivalentes portées par *résultat(s)* ; dans le second cas, la polarité positive du collocatif instancie positivement la bivalence de la base, nous statuons donc sur un mécanisme d'activation positive des potentiels axiologiques portés par *résultat(s)*. Pour *interprétation des résultats*, il s'agit d'un mécanisme de neutralisation.

Au niveau discursif, dans le cas des collocations intrinsèquement bivalentes, on recense deux mécanismes, l'un d'activation positive (avec *améliorer*), l'autre de neutralisation (avec *présenter*).

Activation positive

Ex : *la classification des documents trouvés permet d'améliorer les **résultats de la recherche*** (AR_0011)

Neutralisation

Ex : *cet article présente les **résultats de cette étude*** (AC_0082)

Comme pour les collocations construites avec *évaluation* et un adjectif bivalent, ce mécanisme discursif de neutralisation cotextuelle correspond à un énoncé particulier, situé en introduction de l'article et visant à en présenter la structure. Cette neutralité est ensuite rompue, dans la section consacrée à la présentation des résultats et à leur évaluation justement !

Dans le cas des collocations intrinsèquement axiologisées positivement, l'étude co- et contextuelle fait apparaître trois mécanismes : le renforcement systématiquement positif (avec *permettre d'augmenter*, *validation homogène pour plus de 90 % des familles*, *permettre une réduction substantielle du temps d'intervention manuelle* et *améliorer l'objectivité*), l'activation systématiquement négative (avec *montrer certaines erreurs* et *empêcher*) et enfin la neutralisation (avec *conserver*, avec l'association de *même significatif* et *rester conjectural*, puis *moins tranché* et *seconde vérification*).

Renforcement positif

Ex : *l'augmentation du nombre des analyseurs permet d'augmenter la **fiabilité des résultats*** (AC_0096)

Ex : *l'analyse des **résultats** sur les corpus montre que la validation est homogène pour plus de 90 % des familles* (AC_0040)

Ex : *nous présentons dans cette communication des outils et une méthodologie qui permettent une réduction substantielle du temps d'intervention manuelle et améliorent l'objectivité et la **cohérence du résultat*** (AC_0033)

Activation négative

Ex : *cette très forte condition d'indépendance [...] empêche l'exploitation des résultats théoriques de ces algorithmes semi-supervisés* (AC_0083)

Neutralisation

Ex : *l'UIL est alors conservée dans la partie analyse du résultat* (AR_0014)

Ex : *par ailleurs, l'interprétation des résultats, même significatifs, reste bien entendu conjecturale* (AR_0006)

Ex : *dans les cas où la validité des résultats était moins tranchée, nous avons procédé à une seconde vérification* (AR_0002)

3.2.3.1.2 Le monovalent négatif

Le TSA est appliqué à la collocation *dégradation des résultats*.

Dégradation des résultats

SL – noyau 1 : *dégradation des résultats* → ce qui arrive à la suite et comme effet de qqc. → Donc affaiblissement, diminution, perte de valeur...

L'application du TSA montre que l'adjonction du collocatif *dégradation* à la base *résultat(s)* forme un déploiement argumentatif possible (H32.1). Par conséquent, nous concluons pour *dégradation des résultats* à un mécanisme de motivation argumentative (H32.2).

Nous utilisons le TEA pour statuer sur son axiologisation intrinsèque (H32.3) :

Dégradation des résultats → *Donc amélioration de l'analyse

Dégradation des résultats → Pourtant amélioration de l'analyse

Dégradation des résultats → Donc altération de l'analyse

Dégradation des résultats → *Pourtant altération de l'analyse

Dans ce cas, le collocatif est en position tête et l'argumentation est normative. Seuls les enchaînements transgressif avec « amélioration » et normatif avec « altération » sont sémantiquement acceptables, ce qui signifie que le déploiement argumentatif est intrinsèquement axiologisé négativement. La polarité négative du collocatif contamine donc la bivalence de la base.

Au niveau des mécanismes macro-sémantiques d'axiologisation de la signification (H41), la polarité négative du collocatif contamine donc la bivalence de la base, ce qui nous permet de statuer sur un mécanisme d'activation des valeurs axiologiques bivalentes portées par *résultat(s)*.

Au niveau discursif, l'observation cotextuelle de la collocation fait apparaître un seul mécanisme d'interversion, avec la juxtaposition de *permettre* et *sans*.

Interversion

Ex : *il apparaît que les informations sur les groupes syntaxiques permettent essentiellement de réduire le volume des données de cooccurrences à manipuler sans dégradation des résultats* (AC_0049)

3.2.3.1.3 Les bivalents

L'application du TSA concerne quatre collocations : *résultat du processus*, *résultats des expériences*, *résultats des travaux* et *résultats de qualité*.

Résultat du processus

SL – noyau 1 : *résultat du processus* → ce qui arrive à la suite et comme effet de qqc. → Donc procédure (aboutissant à un résultat)...

Résultats des expériences

SL – noyau 1 : *résultats des expériences* → ce qui arrive à la suite et comme effet de qqc. → Donc provocation d'un phénomène dans l'intention d'en étudier les résultats...

Résultats des travaux

SL – noyau 1 : *résultats des travaux* → ce qui arrive à la suite et comme effet de qqc. → Donc activités exercées pour parvenir à un résultat déterminé...

Résultats du test

SL – noyau 1 : *résultats du test* → ce qui arrive à la suite et comme effet de qqc. → Donc opération permettant de juger, confronter un fait avec une hypothèse...

Qualité des résultats

Les exemples contenant cette collocation montrent que *qualité*, possiblement monovalent positif et bivalent, est employé essentiellement au sens de bivalent.

Ex : *la qualité des résultats produits est bonne (haute précision)* (AR_0009)

Ex : *la qualité des résultats à la sortie du premier module a été évaluée manuellement* (AC_0072)

SL – noyau 1 : *résultat de qualité* → ce qui arrive à la suite et comme effet de qqc. → Donc caractéristique, valeur...

L'application du TSA montre que l'adjonction du collocatif à la base forme avec les stéréotypes proposés ou envisageables de *résultat(s)* un bloc d'argumentation interne (H31.1). Par conséquent, nous concluons pour chacune de ces collocations à la concrétisation d'un PA dans le discours (H31.2). Il s'agit de collocations argumentatives formant un bloc d'argumentation externe. Concernant la collocation *qualité des résultats*, l'application du TSA montre que l'adjonction du collocatif *qualité* à la base *résultat(s)* forme un déploiement argumentatif possible (H32.1). Par conséquent, nous concluons à un mécanisme de motivation argumentative (H32.2).

Nous appliquons le TEA pour statuer sur leur axiologisation intrinsèque :

Résultat du processus → Donc amélioration de l'analyse

Résultat du processus → *Pourtant amélioration de l'analyse

Résultat du processus → Donc altération de l'analyse

Résultat du processus → *Pourtant altération de l'analyse

Qualité des résultats → Donc amélioration de l'analyse

Qualité des résultats → *Pourtant amélioration de l'analyse

Qualité des résultats → Donc altération de l'analyse

Qualité des résultats → *Pourtant altération de l'analyse

Dans cet ensemble de collocations, soit la base occupe la position tête, soit c'est le collocatif. Tous les blocs d'argumentations déployés sont normatifs. Toutefois, cette alternance n'a aucun effet sur l'axiologisation des collocations, car l'application du TEA montre que tous les enchaînements argumentatifs sont sémantiquement acceptables, ce qui signifie que chacune de ces argumentations est intrinsèquement bivalente, en l'absence de tout élément co- ou contextuel.

Au niveau des mécanismes macro-sémantiques d'axiologisation de la signification (H41), la juxtaposition au sein d'une même collocation de deux unités lexicales bivalentes génère une collocation bivalente. Nous statuons donc sur un mécanisme de stabilisation des valeurs axiologiques bivalentes portées par *résultat(s)*.

Au niveau discursif, l'étude des collocations fait apparaître l'existence de trois mécanismes : le renforcement positif (avec la juxtaposition de *garantir* et de *bonne*), l'activation également positive (avec *améliorer les performances*, *information obtenue*), puis la neutralisation (avec *tenir compte*, *présenter* et *discuter*).

Renforcement positif

Ex : *l'approche symbolique garantissant une bonne **qualité des résultats*** (AC_0083)

Activation positive

Ex : *les **résultats des expériences** menées dans cette étude ont montré que l'acquisition de corpus sur le Web pouvait améliorer les performances d'un étiqueteur* (AC_0116)

Ex : *l'information obtenue comme **résultat du second processus** (dans le cas d'une expression déclarative)* (AR_0020)

Neutralisation

Ex : *le modèle tient compte également des **résultats des travaux** précédents [van Santen J. & Hirschberg J. 1994]* (AR_0050)

Ex : *les **résultats du test** sont présentés dans le tableau 3* (AC_0121)

Ex : *la conclusion discute la **qualité des résultats** obtenus et présente les perspectives de ce travail* (AC_0003)

Conformément à ce que nous avons observé précédemment, lorsque le mécanisme de neutralisation accompagne des collocations intrinsèquement bivalentes, les énoncés concernés sont d'ordre « méta-informationnel » (comme pour *résultats des travaux*) ; ils concernent généralement la structure de l'article, sa forme et non sa matière, son fond.

3.2.3.1.4 Les neutres

Le TSA est appliqué aux quatre collocations suivantes : *résultat d'une opération*, *résultat de la méthode*, *résultats du système* et *comparaison des résultats*.

Résultat d'une opération

Dans cette collocation, *opération* est entendu au sens de calcul.

SL – noyau 1 : *résultat d'une opération* → ce qui arrive à la suite et comme effet de qqc. → Donc processus permettant d'engendrer un résultat déterminé...

Résultat de la méthode

SL – noyau 1 : *résultat de la méthode* → ce qui arrive à la suite et comme effet de qqc. → Donc conduite et expression de la pensée... Donc procédé d'investigation, moyen...

Résultats du système

SL – noyau 1 : *résultat du système* → ce qui arrive à la suite et comme effet de qqc. → Donc traitements matériels et logiciels de données...

Comparaison des résultats

SL – noyau 1 : *comparaison des résultats* → ce qui arrive à la suite et comme effet de qqc. → *Donc analyse, jugement, rapprochement...

SL – noyau 1 : *comparaison des résultats* → ce qui arrive à la suite et comme effet de qqc. → *Pourtant analyse, jugement, rapprochement...

Pour *résultat d'une opération*, *résultat de la méthode* et *résultats de cet outil*, l'application du TSA montre que l'adjonction du collocatif à la base forme avec les stéréotypes proposés ou envisageables de *résultat(s)* un bloc d'argumentation interne (H31.1). Par conséquent, nous concluons pour chacune de ces collocations à la concrétisation d'un PA dans le discours (H31.2). Il s'agit de collocations argumentatives formant un bloc d'argumentation externe. Concernant *résultats du système*, l'application du TSA montre que l'adjonction du collocatif à la base forme un déploiement argumentatif possible (H32.1). Par conséquent, nous concluons à un mécanisme de motivation argumentative (H32.2). Enfin, pour *comparaison des résultats*, l'application du TSA montre que l'adjonction du collocatif à la base ne forme ni un bloc d'argumentation interne, ni déploiement argumentatif possible (H33.1). Par conséquent, nous qualifions cette collocation d'arbitraire (H33.2).

Nous appliquons le TEA pour statuer sur leur axiologisation intrinsèque :

Résultat d'une opération → Donc amélioration de l'analyse

Résultat d'une opération → *Pourtant amélioration de l'analyse

Résultat d'une opération → Donc altération de l'analyse

Résultat d'une opération → *Pourtant altération de l'analyse

Résultat du système → Donc amélioration de l'analyse

Résultat du système → *Pourtant amélioration de l'analyse

Résultat du système → Donc altération de l'analyse

Résultat du système → *Pourtant altération de l'analyse

Comparaison des résultats → Donc amélioration de l'analyse

Comparaison des résultats → *Pourtant amélioration de l'analyse

Comparaison des résultats → Donc altération de l'analyse

Comparaison des résultats → *Pourtant altération de l'analyse

Dans cet ensemble de collocations, soit la base occupe la position tête, soit c'est le collocatif, et tous les enchaînements déployés sont normatifs. Toutefois, comme pour les collocatifs bivalents, cette alternance n'a aucun effet sur l'axiologisation des collocations, l'application du TEA montre que tous les enchaînements argumentatifs déployés sont sémantiquement acceptables, ce qui signifie que chacune de ces collocations est intrinsèquement bivalente, en l'absence de tout élément co- ou contextuel. Quant à la collocation comparaison des résultats, elle est également bivalente.

Au niveau des mécanismes macro-sémantiques d'axiologisation de la signification (H41), la présence au sein d'une même collocation d'une unité lexicale bivalente et d'une unité neutre génère une collocation bivalente. Nous statuons donc sur un mécanisme de neutralisation des potentiels axiologiques bivalents. Nous attribuons ce mécanisme à toutes les collocations ci-dessus analysées.

Au niveau discursif, l'étude co- et contextuelle des collocations fait apparaître l'existence de deux mécanismes, l'un de renforcement positif (avec *cumuler les avantages*), l'autre de neutralisation sur des énoncés « méta-informationnels » (avec *voici, évaluer, présenter*).

Renforcement positif

Ex : la *comparaison des résultats* des deux systèmes montre que LOGUS cumule les avantages (AC_0074)

Neutralisation

Ex : en comparaison, voici le *résultat de la méthode* du score d'alignement et des correspondances terminologiques pour les mêmes textes (figure 4) (AC_0021)

Ex : les *résultats du système LIA_PHON* sont présentés dans les tableaux (AR_0042)

3.2.3.2 Le collocatif est un adjectif

3.2.3.2.1 Les monovalents positifs

Les collocations construites avec des adjectifs monovalents positifs sont très nombreuses. Le TSA est appliqué aux collocations *résultat complet*, *résultat correct*, *résultat encourageant*, *résultat intéressant*, *résultat positif*, *résultat satisfaisant*, *résultats bons*, *résultats meilleurs*, *résultats pertinents*, *résultats prometteurs*, *résultats significatifs*, *résultats supérieurs*, *excellents résultats*.

Résultat complet

SL – noyau 1 : *résultat complet* → ce qui arrive à la suite et comme effet de qqc. → Donc avec toutes les parties, entier, total, achevé...

Résultat correct

SL – noyau 1 : *résultat correct* → ce qui arrive à la suite et comme effet de qqc. → Donc bon, exact, juste, possible...

Résultat encourageant

SL – noyau 1 : *résultat encourageant* → ce qui arrive à la suite et comme effet de qqc. → Donc favorable, prometteur, stimulant...

Résultat intéressant

SL – noyau 1 : *résultat intéressant* → ce qui arrive à la suite et comme effet de qqc. → Donc captivant... Donc important... Donc avantageux, valable...

Résultat positif

SL – noyau 1 : *résultat positif* → ce qui arrive à la suite et comme effet de qqc. → Donc appréciation, jugement de valeur favorable...

Résultat satisfaisant

SL – noyau 1 : *résultat satisfaisant* → ce qui arrive à la suite et comme effet de qqc. → Donc bon, convenable...

Résultats bons

SL – noyau 1 : *résultats bons* → ce qui arrive à la suite et comme effet de qqc. → Donc positif, utile, agréable, efficace...

Résultats meilleurs

SL – noyau 1 : *résultats meilleurs* → ce qui arrive à la suite et comme effet de qqc. → Donc de qualité supérieure...

Résultats pertinents

SL – noyau 1 : *résultats pertinents* → ce qui arrive à la suite et comme effet de qqc. → Donc adapté... Donc à-propos, juste...

Résultats prometteurs

SL – noyau 1 : *résultats prometteurs* → ce qui arrive à la suite et comme effet de qqc. → Donc assurance, espérance... Donc encourageant, engageant...

Résultats significatifs

SL – noyau 1 : *résultats significatifs* → ce qui arrive à la suite et comme effet de qqc. → Donc preuve, confirmation...

Résultats supérieurs

SL – noyau 1 : *résultats supérieurs* → ce qui arrive à la suite et comme effet de qqc. → Donc plus grand, plus élevé... Donc meilleur...

Excellents résultats

SL – noyau 1 : *excellents résultats* → ce qui arrive à la suite et comme effet de qqc. → Donc très bon, de parfaite qualité...

L'application du TSA montre que l'adjonction du collocatif à la base *résultat(s)* forme un déploiement argumentatif possible (H32.1). Par conséquent, nous concluons pour chacune de ces collocations à un mécanisme de motivation argumentative (H32.2).

Pour statuer sur l'axiologisation intrinsèque, nous appliquons le TEA (H32.3) :

Résultat correct → Donc amélioration de l'analyse

Résultat correct → *Pourquoi amélioration de l'analyse

Résultat correct → *Donc altération de l'analyse

Résultat correct → Pourquoi altération de l'analyse

Excellents résultats → Donc amélioration de l'analyse

Excellents résultats → *Pourquoi amélioration de l'analyse

Excellents résultats → *Donc altération de l'analyse

Excellents résultats → Pourquoi altération de l'analyse

La base est toujours en position tête, sauf pour la collocation *excellents résultats*, et l'argumentation déployée est toujours normative. Seuls les enchaînements argumentatifs normatif avec « amélioration » et transgressif avec « altération » sont sémantiquement acceptables, ce qui signifie que les collocations sont intrinsèquement axiologisées positivement. La polarité positive du collocatif semble contaminer la base.

Au niveau des mécanismes macro-sémantiques d'axiologisation de la signification (H41), nous statuons sur un mécanisme d'activation positive des valeurs axiologiques bivalentes portées par *résultat(s)*. Nous attribuons ce mécanisme à toutes les collocations ci-dessus.

Au niveau discursif, l'étude co- et contextuelle des collocations montre que tous les mécanismes sont possibles : le renforcement (avec l'association de *donner*, de *comme les trois précédentes* et *très*, puis de *permettre* et *traiter les problèmes*), l'activation (avec *permettre d'obtenir*, *donner* et *obtenir*), la neutralisation (avec *représenter*, *consister* et *montrer*), l'affaiblissement (avec *il n'est pas exclu*, *raisonnable de supposer*, *pouvoir être améliorés* et l'association de *garantir* et de *forcer*) et enfin l'interversion (avec *réduire considérablement*).

Renforcement

Ex : *cette dernière expérience donne donc, comme les trois précédentes, de très **bons résultats*** (AR_0060)

Ex : *ces **excellents résultats** nous permettent de traiter les problèmes liés à la capitalisation de textes présentés précédemment* (AR_0042)

Activation

Ex : *elle a permis d'obtenir des **résultats encourageants**, tant pour la qualité des ressources acquises que pour la quantité de temps passé* (AC_0118)

Ex : *elle donne aussi des **résultats intéressants** d'un point de vue sémantique* (AC_0057)

Ex : *il faut prouver que la procédure de construction donne le **bon résultat*** (AR_0019)

Ex : *nous reportons en table 2 ces taux sur nos bitextes pour le SMT obtenant les **meilleurs résultats*** (AC_0043)

Ex : *cela était non seulement nécessaire pour obtenir des **résultats pertinents** sur la validité de notre modèle, mais aussi (et surtout) pour rendre la ressource annotée réutilisable* (AR_0003)

Ex : *en se basant sur $n=4$ ont obtenu des **résultats significatifs*** (AC_0008)

Ex : *cette méthode (PLI avec expert par la suite) donne d'**excellents résultats** pour l'extraction de couples qualia* (AC_0083)

Neutralisation

Ex : *le **résultat complet** de l'évaluation d'un système d'étiquetage est représenté sous la forme de 4 points [...] dans la portion de plan $[0,1] \times [0,1]$* (AC_0023)

Ex : *le **résultat correct** ou clé, consiste ici en quatre ensembles ou classes* (AR_0053)

Ex : *le traitement d'une langue proche de la langue générale ici tenté avec des outils destinés à des langues techniques a montré un certain nombre de **résultats positifs*** (AR_0069)

Affaiblissement

Ex : *il est raisonnable de supposer que l'amélioration de F_c peut conduire à des **résultats meilleurs** pour F_a* (AC_0008)

Ex : *nous pensons que ces **résultats prometteurs** peuvent être améliorés par un raffinement de cette approximation* (AC_0008)

Ex : *nous admettons que l'on garantit l'obtention d'un **résultat satisfaisant** en forçant les règles de réduction à s'appliquer de manière descendante* (AR_0019)

Ex : *il n'est pas exclu qu'une combinaison de classifieurs produise des **résultats supérieurs** à ceux produits par des classifieurs travaillant individuellement* (AC_0105)

Interversion

Ex : *l'alignement d'un texte plus linéaire, comme un roman, réduit considérablement la **pertinence des résultats*** (AC_0021)

3.2.3.2.2 Les monovalents négatifs

Le TSA est appliqué aux collocations *résultat décevant* et *résultats mauvais*.

Résultat décevant

SL – noyau 1 : *résultat décevant* → ce qui arrive à la suite et comme effet de qqc. → Donc ne satisfait pas à l'attente... Donc décourageant...

Résultats mauvais

SL – noyau 1 : *résultats mauvais* → ce qui arrive à la suite et comme effet de qqc. → Donc défaut, imperfection essentielle...

L'application du TSA montre que l'adjonction du collocatif à la base *résultat(s)* forme un déploiement argumentatif possible (H32.1). Nous concluons pour ces deux collocations à un mécanisme de motivation argumentative (H32.2).

Nous utilisons le TEA pour statuer sur leur axiologisation intrinsèque (H32.3) :

Résultat décevant → *Donc amélioration de l'analyse

Résultat décevant → Pourtant amélioration de l'analyse

Résultat décevant → Donc altération de l'analyse

Résultat décevant → *Pourtant altération de l'analyse

Dans ces deux cas, le collocatif est en position tête (toutefois, *mauvais résultat* existe aussi) et l'argumentation est normative. Seuls les enchaînements transgressif avec « amélioration » et normatif avec « altération » sont sémantiquement acceptables, ce qui signifie que le déploiement argumentatif est intrinsèquement axiologisé négativement.

Au niveau des mécanismes macro-sémantiques d'axiologisation de la signification (H41), la polarité négative du collocatif contamine donc la bivalence de la base, ce qui nous permet de statuer sur un mécanisme d'activation négative des valeurs axiologiques bivalentes portées par *résultat*.

Au niveau discursif, l'observation cotextuelle des deux collocations fait apparaître un seul mécanisme d'activation de la polarité négative (avec *conduire* et *donner*).

Activation

Ex : *les expériences que nous avons menées dans cette voie ont conduit à des **résultats décevants*** (AC_0085)

Ex : *ce qui [...] donne de **mauvais résultats** dans l'interprétation ultérieure par ha22* (AR_0071)

3.2.3.2.3 Les bivalents

Le TSA est appliqué aux deux collocations suivantes : *résultats expérimentaux* et *résultats préliminaires*.

Résultats expérimentaux

SL – noyau 1 : *résultats expérimentaux* → ce qui arrive à la suite et comme effet de qqc. → Donc étude d'un phénomène provoqué pour observer les résultats...

Résultats préliminaires

SL – noyau 1 : *résultats préliminaires* → ce qui arrive à la suite et comme effet de qqc. → *Donc préalable, préparatoire (à qqc. de plus important)...

SL – noyau 1 : *résultats préliminaires* → ce qui arrive à la suite et comme effet de qqc. → *Pourtant préalable, préparatoire (à qqc. de plus important)...

L'application du TSA montre que l'adjonction du collocatif *expérimentaux* à la base forme avec les stéréotypes proposés ou envisageables de *résultats* un bloc d'argumentation interne (H31.1). Par conséquent, nous concluons à la concrétisation d'un PA dans le discours (H31.2). Il s'agit d'une collocation argumentative. Dans le cas de *résultats préliminaires*, l'application du TSA montre que l'adjonction du collocatif à la base ne forme ni un bloc d'argumentation interne, ni un déploiement argumentatif possible (H33.1). Par conséquent, nous concluons à une collocation arbitraire (H33.2).

Nous appliquons le TEA pour statuer sur leur axiologisation intrinsèque :

Résultats expérimentaux → Donc améliorer l'analyse

Résultats expérimentaux → *Pourtant améliorer l'analyse

Résultats expérimentaux → Donc altérer l'analyse

Résultats expérimentaux → *Pourtant altérer l'analyse

Résultats préliminaires → Donc améliorer l'analyse

Résultats préliminaires → *Pourtant améliorer l'analyse

Résultats préliminaires → Donc altérer l'analyse

Résultats préliminaires → *Pourtant altérer l'analyse

Dans cet ensemble de collocations, la base occupe systématiquement la position tête et les blocs d'argumentations déployés sont normatifs. L'application du TEA montre que tous les enchaînements argumentatifs sont sémantiquement acceptables, ce qui signifie que ces collocations conservent leur potentiel axiologique bivalent, en l'absence de tout élément co- ou contextuel.

Au niveau des mécanismes macro-sémantiques d'axiologisation de la signification (H41), la juxtaposition au sein d'une même collocation de deux unités lexicales bivalentes génère une collocation bivalente. Comme pour les collocations construites avec un nom, nous statuons sur un mécanisme macro-sémantique dit de stabilisation des valeurs axiologiques bivalentes portées par *résultats*.

Au niveau discursif, l'étude co- et contextuelle des collocations fait apparaître l'existence de deux mécanismes possibles, l'un de renforcement positif (avec *obtenir* accolé à *mettre en évidence l'intérêt*, puis *obtenir* et *encourageant*) et négatif (avec *infirmer la validité*), l'autre de neutralisation (avec *montrer, présenter*).

Renforcement positif

Ex : la Section 4 présente quelques **résultats expérimentaux** obtenus qui permettent de mettre en évidence l'intérêt de notre approche (AC_0087)

Ex : les **résultats préliminaires** obtenus sont encourageants quant à la pertinence de ces deux principes (AC_0022)

Renforcement négatif

Ex : des **résultats expérimentaux** infirment la validité du principe RA pour l'espagnol (AC_0022)

Neutralisation

Ex : nous montrons des *résultats expérimentaux* pour la lettre e sur le MeSH français (AC_0077)

Ex : nous présentons des *résultats préliminaires* d'une méthode de description de la sémantique des formes prédicatives (AC_0004)

3.2.3.2.4 Les neutres

Le TSA est appliqué aux collocations *résultat empirique*, *résultat final*, *résultat global*, *résultat théorique*, *résultats numériques*, *résultats quantitatifs* et *résultats statistiques*.

Résultat empirique

SL – noyau 1 : *résultat empirique* → ce qui arrive à la suite et comme effet de qqc. → Donc qui résulte de l'expérience... Donc expérimental, pragmatique, pratique...

Résultat final

SL – noyau 1 : *résultat final* → ce qui arrive à la suite et comme effet de qqc. → Donc définitif, terminal, ultime... Donc terme, but... Renforcement

Résultat global

SL – noyau 1 : *résultat global* → ce qui arrive à la suite et comme effet de qqc. → Donc considéré dans son ensemble... Donc complet, intégral...

Résultat théorique

SL – noyau 1 : *résultat théorique* → ce qui arrive à la suite et comme effet de qqc. → *Donc connaissance abstraite, spéculation...

SL – noyau 1 : *résultat théorique* → ce qui arrive à la suite et comme effet de qqc. → *Pourtant connaissance abstraite, spéculation...

Résultats numériques

SL – noyau 1 : *résultats numériques* → ce qui arrive à la suite et comme effet de qqc. → *Donc évalué en nombre, arithmétique...

SL – noyau 1 : *résultats numériques* → ce qui arrive à la suite et comme effet de qqc. → *Pourtant évalué en nombre, arithmétique...

Résultats quantitatifs

SL – noyau 1 : *résultats quantitatifs* → ce qui arrive à la suite et comme effet de qqc. → *Donc (concerne) un nombre, une mesure, un poids...

SL – noyau 1 : *résultats quantitatifs* → ce qui arrive à la suite et comme effet de qqc. → *Pourtant (concerne) un nombre, une mesure, un poids...

Résultats statistiques

SL – noyau 1 : *résultats statistiques* → ce qui arrive à la suite et comme effet de qqc. → *Donc données numériques...

SL – noyau 1 : *résultats statistiques* → ce qui arrive à la suite et comme effet de qqc. → *Pourtant données numériques...

Pour *résultat empirique*, *résultat final* et *résultat global*, l'application du TSA montre que l'adjonction du collocatif à la base forme avec les stéréotypes proposés ou envisageables de *résultat(s)* un bloc d'argumentation interne (H31.1). Par conséquent, nous concluons pour chacune de ces collocations à la concrétisation d'un PA dans le discours (H31.2). Il s'agit de collocations argumentatives formant un bloc d'argumentation externe. Pour les autres collocations, l'application du TSA montre que l'adjonction du collocatif à la base ne forme ni un bloc d'argumentation interne, ni un déploiement argumentatif possible (H33.1). Par conséquent, nous concluons pour chacune de ces collocations à une collocation arbitraire (H33.2).

Nous appliquons le TEA pour statuer sur leur axiologisation intrinsèque :

Résultat empirique → Donc améliorer l'analyse

Résultat empirique → *Pourtant améliorer l'analyse

Résultat empirique → Donc altérer l'analyse

Résultat empirique → *Pourtant altérer l'analyse

Résultat théorique → Donc améliorer l'analyse

Résultat théorique → *Pourtant améliorer l'analyse

Résultat théorique → Donc altérer l'analyse

Résultat théorique → *Pourtant altérer l'analyse

Dans ce paradigme collocationnel, la base occupe la position tête de la collocation et tous les blocs d'argumentations déployés sont normatifs. L'application du TEA montre que tous les enchaînements argumentatifs sont sémantiquement acceptables, ce qui signifie que ces collocations sont axiologiquement bivalentes. Pour les collocations arbitraires, l'analyse aboutie à la même conclusion.

Au niveau des mécanismes macro-sémantiques d'axiologisation de la signification (H41), la présence au sein d'une même collocation d'une unité lexicale bivalente et d'une unité neutre génère une collocation bivalente. Nous statuons donc sur un mécanisme de neutralisation des potentiels axiologiques bivalents.

Au niveau discursif, l'étude des collocations fait apparaître la possibilité de trois mécanismes : le renforcement positif (avec *très satisfaisant*), l'activation positive (avec *sans altérer*, *déterminant pour la qualité*, *permettre une analyse complémentaire*, *connaître*, *permettre d'uniformiser*, *donner*, *important*) ou négative (avec *sans doute améliorer*, *rester* et *étude plus fine nécessaire*) et la neutralisation (avec *présenter*).

Renforcement positif

Ex : ce premier **résultat empirique** est très satisfaisant (AC_0038)

Activation positive

Ex : sans que son intervention n'altère le **résultat final** (AR_0071)

Ex : les mots-clés initiaux [...] sont déterminants pour la qualité du **résultat final** (AC_0068)

Ex : les mesures effectuées au niveau local permettent une analyse complémentaire des **résultats globaux** et sont plus orientées vers le diagnostic (AR_0054)

Ex : nous désirions en effet connaître les **résultats réels** de notre système plutôt que le résultat théorique lié exclusivement à la désambiguïisation lexicale sémantique (AR_0058).

Ex : enfin, de récents **résultats théoriques** sur la tabulation vont permettre d'uniformiser le traitement des formalismes actuels (AC_0075)

Ex : nous donnons quelques exemples et des **résultats numériques** obtenus (AC_0053)

Ex : la monotonie des **résultats statistiques** est importante (AR_0071)

Activation négative

Ex : on peut sans doute améliorer les **résultats globaux** en combinant l'identité (les transfuges) et la ressemblance (les *n*-grammes ou autres) (AR_0062)

Ex : les résultats présentés restent des **résultats quantitatifs globaux** et une étude qualitative plus fine serait nécessaire (AC_0049)

Neutralisation

Ex : nous présentons ses **résultats empiriques** (AR_0012)

Ex : nous présentons quelques **résultats numériques** supplémentaires (AR_0065)

Là encore, il s'agit d'énoncés « méta-informationnels ».

3.3 Synthèse sur les bivalents

De même que pour les monovalents positifs et négatifs, cette synthèse repose sur les analyses sémantico-discursives menées, concernant le binôme bivalent *évaluer-évaluation(s) – résultat(s)*. Ces deux bases rassemblent à elles deux 83 collocations différentes et 845 occurrences.

Les bivalents ont généré énormément de collocations argumentatives, parmi lesquelles *évaluation de la qualité, critère(s) d'évaluation, évaluation des algorithmes, résultat du processus, résultats d'une opération*, etc. Pour toutes ces collocations, l'application du TSA a montré que l'adjonction du collocatif à la base forme avec les stéréotypes proposés ou envisageables de *évaluer-évaluation(s)* ou *résultat(s)* un bloc d'argumentation interne (H31.1). Aussi, toutes ces collocations correspondent à la concrétisation d'un PA dans le discours (H31.2) ; elles sont prévisibles, « calculables » sur la base des PA existant pour chaque mot. Là encore, nous préférons les caractériser de bloc d'argumentation externe sous-jacent.

Conformément aux postulats théoriques de la SPA-ALD, de par le lien argumentatif posé, les PA sont par essence évaluatifs. Le TEA a montré que toutes les collocations argumentatives relevées sont axiologisées soit positivement, comme pour *évaluation fine*, soit elles restent intrinsèquement bivalentes, comme pour *évaluation comparative*.

Le PA est axiologisé positivement lorsque le collocatif adjectival, postposé par rapport à la base, est positif. Dans ce cas, le mécanisme macro-sémantique correspond à une activation positive de la bivalence de la base par le collocatif, comme pour *évaluation objective*.

Le PA est bivalent lorsque le collocatif nominal, antéposé ou postposé par rapport à la base, est bivalent ou neutre. Dans ce cas, l'antéposition ou la postposition du collocatif par rapport à la base n'influe pas sur l'axiologisation de l'ensemble de la collocation. La collocation demeure bivalente, créant un mécanisme de neutralisation (stabilisation) (H41), comme pour *résultats d'expérience* ou *méthodologie(s) d'évaluation* par exemple.

Conformément aux postulats SPA-ALD, les mécanismes discursifs des bivalents correspondent essentiellement à des phénomènes de renforcement ou d'activation marqués axiologiquement (positivement ou négativement) ; nous appliquons les mêmes mécanismes au niveau macro-sémantique. Concernant ces collocations argumentatives, nous avons identifié les deux mécanismes, dans leur forme positive et négative :

- renforcement positif : **Ex** : une évaluation automatique *peut être équivalente à une évaluation d'experts* ;
- renforcement négatif : **Ex** : il est *toujours difficile* d'évaluer la qualité d'un étiquetage ;
- activation positive : **Ex** : évaluation des performances *obtenues* ; *s'accorder sur* une mesure d'évaluation ;
- activation négative : **Ex** : *il n'est donc pas possible de faire* un programme d'évaluation ; une évaluation automatique *peut aussi ne pas toujours être souhaitable*.

Les bivalents ont généré une très grande quantité de collocations argumentativement motivées, parmi lesquelles *évaluation de la cohésion*, *évaluation des réponses*, *évaluation des résultats*, *difficile à évaluer*, *résultats de la recherche*, etc. Pour toutes les collocations concernées, l'application du TSA a montré que l'adjonction du collocatif à la base *évaluer-évaluation(s)* ou *résultat(s)* forme un déploiement argumentatif possible (H32.1). Aussi, toutes ces collocations correspondent à un mécanisme de motivation argumentative (H32.2). Elles entretiennent une relation argumentative avec la base. Le mécanisme de motivation argumentative revêt un caractère moins prédictible que celui mobilisé par les collocations argumentatives, sans que l'émergence de régularités en fonction de la monovalence négative de la base soit compromise. De même que pour les collocations argumentatives, nous trouvons plus juste de parler de réalisation de surface d'un bloc d'argumentation externe sous-jacent.

Conformément aux postulats théoriques de la SPA-ALD, de par le lien argumentatif déployé, ces collocations comportent nécessairement une évaluation, orientée vers l'un ou l'autre des pôles axiologiques, positif ou négatif (H32.3). Le TEA a montré que pour tous les cas de motivation argumentative, l'orientation axiologique des collocations est soit positive, comme pour *résultat correct*, soit négative, comme pour *coût de l'évaluation*, soit bivalente, comme pour *outils d'évaluation*.

La collocation est axiologisée positivement lorsque le collocatif, qu'il s'agisse d'un nom ou d'un adjectif antéposé ou postposé par rapport à la base, est positif. Dans ce cas, le mécanisme macro-sémantique est unique, il s'agit d'une activation positive de la collocation, comme pour *évaluation qualitative*, *cohérence du résultat*, *fiabilité du résultat*.

La collocation est axiologisée négativement lorsque le collocatif nominal, antéposé ou postposé par rapport à la base, est négatif. Dans ce cas, le mécanisme macro-sémantique est unique, il s'agit d'une activation négative de la collocation, comme pour *coût de l'évaluation*, *dégradation des résultats*.

La collocation est bivalente lorsque le collocatif nominal, antéposé ou postposé par rapport à la base, est positif, bivalent ou neutre. Les mécanismes observés sont les suivants :

- quand le collocatif positif est postposé, il s'agit d'un mécanisme macro-sémantique de neutralisation, comme pour *résultats de la recherche* ;
- quand le collocatif bivalent est antéposé ou postposé, il s'agit également d'un mécanisme de neutralisation (stabilisation), comme pour *évaluation globale* ;

- quand le collocatif neutre est antéposé ou postposé, il s'agit également d'un mécanisme de neutralisation, comme pour *évaluation des méthodes, résultats d'un système*.

Concernant les mécanismes discursifs généraux, ils varient selon que la collocation est axiologisée positivement-négativement ou selon qu'elle est bivalente :

- quand la collocation est axiologisée, deux mécanismes ont été rencontrés : la neutralisation (**Ex** : *chercher à évaluer la cohésion*) et l'interversion (**Ex** : *manipuler sans dégradation des résultats*) ;
- quand la collocation est bivalente, trois mécanismes ont été trouvés : le renforcement positif (**Ex** : *une énumération fastidieuse et difficile à évaluer*), l'activation positive (**Ex** : *la classification des documents trouvés permet d'améliorer les résultats*) et l'activation négative (**Ex** : *les problèmes mis en évidence dans les phases d'évaluation*).

Enfin, nous avons trouvé quatre collocations arbitraires formées sur les bases *évaluation(s)* et *résultat(s)* : *évaluation manuelle*, *évaluation automatique*, *interprétation des résultats* et *comparaison des résultats*. La collocation arbitraire correspond au cas où la collocation repérée ne correspond ni à la concrétisation d'un PA dans le discours, ni à un cas de motivation argumentative (H33.1). Dans ce cas, la collocation est qualifiée d'arbitraire (H33.2), au sens où elle ne peut être sémantiquement prédictible ; elle n'a de réalité que discursive.

Par l'absence de lien argumentatif, la collocation arbitraire ne devrait pas être intrinsèquement axiologisée, ce que nous remettons en cause. Le TEA a montré que les collocations étaient systématiquement bivalentes (H33.3), que le collocatif soit positif, bivalent ou neutre. Figurent ci-dessous les mécanismes observés :

- quand le collocatif positif est postposé, il s'agit d'un mécanisme de neutralisation, comme pour *interprétation des résultats* ;
- quand le collocatif bivalent est postposé, il s'agit d'un mécanisme de neutralisation, comme pour *évaluation manuelle* ;
- quand le collocatif neutre est antéposé ou postposé, il s'agit d'un mécanisme de neutralisation, comme pour *évaluation automatique* et *comparaison des résultats* ;

Un seul mécanisme discursif d'activation a été rencontré, tantôt positif (**Ex** : *réaliser une évaluation automatique de la pertinence des résultats*), tantôt négatif (**Ex** : *quelque chose reste moins coûteux que l'évaluation manuelle*).

Contrairement aux monovalents négatifs, la confrontation des résultats obtenus pour les trois types de collocation sur le binôme bivalent, révèle une dichotomie systématique, bien que peu surprenante. Chaque fois que la collocation est intrinsèquement axiologisée, les mécanismes constatés sont de l'ordre du renforcement, de la neutralisation ou de l'interversion ; en

revanche chaque fois que la collocation est bivalente, les mécanismes récurrents sont le renforcement positif ou négatif et l'activation positive ou négative.

De même que pour les synthèses précédentes, aucune tendance ne se dégage pour statuer sur l'hypothèse H42 relative à la prépondérance des mécanismes discursifs de renforcement et d'activation dans le discours scientifique spécialisé.

4 Synthèse générale

À l'inverse des synthèses précédentes, centrées sur la valeur axiologique (positive, négative ou bivalente) des binômes retenus, cette synthèse repose sur les régularités observées à l'égard des types de collocations identifiés. Au total, nous avons analysé 6 bases, qui ont généré 164 collocations différentes, rassemblant 1 670 occurrences (*cf.* Tableau 33).

BASE	NOMBRE DE COLLOCATIONS	NOMBRE D'OCCURRENCES
Améliorer-Amélioration(s)	12	217
Performance(s)	21	200
Erreur(s)	16	113
Problème(s)	32	295
Évaluer-Évaluation(s)	37	388
Résultat(s)	46	457
TOTAL	TOTAL	TOTAL
6	164	1670

Tableau 33 : Nombre de collocations et d'occurrences recensées par base

Rappelons d'abord que, par la présence effective de collocations argumentatives, de collocations argumentativement motivées et de collocations arbitraires, notre hypothèse centrale a été validée. Ce faisant, les analyses menées ont révélé un parallèle aussi inattendu qu'intéressant entre la typologie sémantique adoptée pour classer les collocations textuelles et la typologie argumentative obtenue par la validation de notre hypothèse centrale. En effet, il est apparu que les collocations libres, définies comme des associations de mots dont les deux éléments conservent leur sens littéral et dont la combinatoire est assez libre pour sembler quasi illimitée, correspondent aux collocations motivées argumentativement, comme *amélioration sensible*, *problème crucial* ou encore *fiabilité du résultat*. De même, les collocations restreintes, au sens d'associations de mots dont l'un des constituants conserve son sens littéral tandis que l'autre a un sens figuré, se rapprochent des collocations argumentatives, comme *bonne performance*, *problème épineux* ou encore *évaluation fine*. Quant aux collocations arbitraires comme *amélioration des systèmes*, *pourcentage d'erreur*, *problèmes linguistiques* ou encore *évaluation automatique*, elles présentent la particularité d'être représentatives des pratiques et des préoccupations liées au domaine TAL, instanciant ainsi pleinement leur fonction cohésive au sein du corpus représentatif de la communauté de discours TAL.

Concernant les six mécanismes d'axiologisation du sens au niveau discursif général cotextuel et/ou contextuel, nous avons pu observer qu'ils étaient effectivement transposables au niveau macro-sémantique. Les analyses menées ont fait apparaître des mécanismes récurrents, comparables à ceux dégagés par les études menées sur la motivation sémantique. En particulier, les collocations textuelles étant peu imagées, les régularités argumentatives observées correspondent aux mécanismes dits de « convergence de polarité » et de « renforcement », inhérent au principe naturel de « congruence sémantique ». Tel qu'il est décrit dans le cas de la motivation sémantique, le mécanisme de convergence de polarité correspond au cas où une base marquée d'un trait « agréable » sélectionne préférentiellement un collocatif neutre ou également marqué d'un trait « agréable » (**Ex** : *réelle joie* ; *joie sincère*). À l'inverse, lorsqu'une base comporte le trait « désagréable », elle s'adjoint préférentiellement un collocatif neutre ou également marqué d'un trait « désagréable » (**Ex** : *grande peur* ; *peur terrible*). De même, tel qu'il est décrit dans le cas de la motivation sémantique, le mécanisme de renforcement correspond soit au cas où un collocatif négatif est adjoint à une base également négative (**Ex** : *affreuse tristesse*), soit au cas où un collocatif comportant le trait « intensif » est adjoint à une base comportant également le trait « intensif » (**Ex** : *profonde tristesse*).

Concernant nos données, que l'orientation positive ou négative de la collocation résulte d'un mécanisme sémantico-discursif d'axiologisation du sens comme pour les collocations argumentatives ou bien d'un mécanisme pragmatico-discursif comme pour les cas de motivation argumentative et de collocations arbitraires, les quatre mêmes mécanismes macro-sémantiques ont été recensés :

- un mécanisme de renforcement. Ce mécanisme s'applique uniquement dans le cas où la base et le collocatif ont la même orientation axiologique (**Ex** : *gain de performance* ; *problème épineux*) ;
- un mécanisme d'activation. Ce mécanisme s'applique dans deux cas de figure, soit quand un collocatif bivalent ou neutre est adjoint à une base axiologisée (**Ex** : *problème du choix*), soit quand un collocatif axiologisé est adjoint à une base bivalente (**Ex** : *évaluation objective*) ;
- un mécanisme de neutralisation-stabilisation. Ce mécanisme s'applique dans le cas où un collocatif bivalent ou neutre est adjoint à une base bivalente (**Ex** : *résultats d'expériences*) ;
- un mécanisme d'interversion. Ce mécanisme s'applique dans le cas où le collocatif, antéposé par rapport à la base, a une valeur axiologique opposée à celle de la base (**Ex** : *dégradation de performance* ; *solution du problème*).

Toutefois, quelques cas de processus axiologiques résistent à la logique d'un mécanisme en particulier, comme pour *légère amélioration*, *problème classique* ou *résultats de la recherche*, par exemple. Les éléments de réponse semblent s'orienter vers la nature grammaticale du collocatif, son antéposition ou sa postposition par rapport à la base et la nature normative ou transgressive du déploiement argumentatif qui sous-tend l'articulation de la collocation. Pour les mêmes raisons, l'hypothèse posée en cours d'analyse concernant la valeur axiologique de la base supposée plus forte que celle du collocatif est impossible à valider ; la multitude de combinaisons possibles entre ces différents paramètres empêchant de dégager des régularités. Les données analysées sont trop éparses pour statuer sur un mécanisme stable.

D'un point de vue discursif général, nous pensions intuitivement trouver une majorité de potentiels activés, voire renforcés, en particulier pour les collocations arbitraires, mais pas ou peu de potentiels neutralisés ou affaiblis et encore moins de potentiels intervertis ou transgressés, eu égard à un contrat de transparence et de non-ambiguïté dans la transmission du savoir. Or, de même que pour l'hypothèse précédente posée en cours d'analyse, nous avons été surpris par l'absence de régularité et, par répercussion, par la multitude de combinaisons possibles entre les mécanismes discursifs rencontrés, la valeur axiologique de la base et la nature de la collocation. Là encore, les données analysées sont trop éparses pour statuer sur un mécanisme stable. Toutefois, deux éléments sont notoires : d'une part, la présence de tous les mécanismes discursifs possibles, à l'exception du mécanisme de transgression et, d'autre part, la régularité des mécanismes appliqués aux collocations bivalentes, tout type confondu. En effet, concernant les collocations bivalentes, les mécanismes se réduisent au renforcement positif (**Ex : une évaluation automatique peut être équivalente à une évaluation d'experts**) ou négatif (**Ex : il est toujours difficile d'évaluer la qualité d'un étiquetage**) et à l'activation positive (**Ex : évaluation des performances obtenues ; s'accorder sur une mesure d'évaluation**) ou négative (**Ex : une évaluation automatique peut aussi ne pas toujours être souhaitable**).

L'hypothèse générale – H2 – avait été qualifiée de générale, étant donné qu'elle préexistait à l'observation des données du corpus, en qualité de fondement théorique de la SPA. De par le lien argumentatif posé, les PA sont par essence évaluatifs, orientés vers l'un ou l'autre des pôles axiologiques, positif ou négatif, en fonction de la contamination discursive, due soit à l'environnement sémantique, soit à l'environnement cotextuel et/ou contextuel. En ce sens, les PA sont des entités fondamentalement subjectives et le postulat de la SPA établit que la subjectivité du DSS est inscrite dans les PA des unités lexicales qui composent tout discours. Par la présence de collocations argumentatives intrinsèquement et discursivement axiologisées dans le corpus TAL(N), nos analyses corroborent ce postulat. Par répercussion, nous pouvons également valider l'hypothèse H1, concernant la subjectivité réelle du discours scientifique spécialisé. Nous avons qualifié cette hypothèse de volontairement triviale, puisque le fait d'y répondre par l'affirmative présente assez peu d'intérêt, ce qui importe étant de savoir à quel niveau cette subjectivité s'inscrit et à quels mécanismes elle obéit.

Conclusion

Nous avons présenté dans ce chapitre l'ensemble des analyses réalisées sur les six unités lexicales sélectionnées pour former les bases des collocations étudiées, selon leur axiologisation : les monovalents positifs *amélioration* et *performance*, les monovalents négatifs *erreur* et *problème* et les bivalents *évaluation* et *résultat*. Ce faisant, par l'application du protocole de validation aux 6 bases sélectionnées, aux 164 collocations analysées et aux 1 670 occurrences traitées, nous avons pu statuer sur les hypothèses posées et mettre à jour quelques régularités. Tout d'abord, par la présence effective de collocations argumentatives, de collocations argumentativement motivées et de collocations arbitraires, notre hypothèse centrale a été validée. Ce faisant, les analyses menées ont révélé un parallèle intéressant concernant la superposition possible de la typologie sémantique, collocations libres *vs* collocations restreintes adoptée pour classer les collocations textuelles et la typologie argumentative obtenue par la validation de notre hypothèse centrale. En effet, il est apparu

que les collocations libres correspondent aux cas de motivations argumentatives décrits dans notre étude, de même que les collocations restreintes se rapprochent des collocations argumentatives. Quant aux collocations arbitraires, elles présentent la particularité d'être représentatives des pratiques liées au domaine TAL, instanciant ainsi pleinement leur fonction cohésive au sein du corpus représentatif de la communauté de discours TAL. De même, l'hypothèse d'une transposition possible des mécanismes d'axiologisation du sens au niveau discursif général cotextuel et/ou contextuel au niveau macro-sémantique s'est révélée concluante. Les analyses menées ont fait apparaître des mécanismes récurrents, comparables à ceux dégagés par les études menées sur la motivation sémantique. En particulier, les collocations textuelles étant peu imagées, les régularités argumentatives observées correspondent aux mécanismes dits de « convergence de polarité » et de « renforcement », inhérent au principe naturel de « congruence sémantique ». Quels que soient les mécanismes sémantico-discursifs ou pragmatico-discursifs mobilisés par les collocations argumentatives, motivées ou arbitraires, quatre mécanismes macro-sémantiques sont récurrents et suffisent à expliquer les trois quarts des données recensées ; il s'agit des mécanismes de renforcement, d'activation, de neutralisation (stabilisation) et d'interversion. En revanche, d'un point de vue discursif général, les résultats des analyses sont moins probants. Nous pensions intuitivement trouver une majorité de potentiels activés, voire renforcés, en particulier pour les collocations arbitraires, mais pas ou peu de potentiels neutralisés ou affaiblis et encore moins de potentiels intervertis ou transgressés, eu égard à un contrat de transparence et de non-ambiguïté dans la transmission du savoir. Or nous avons été surpris par l'absence de régularité et, *a contrario*, par la multitude de combinaisons possibles entre les mécanismes discursifs rencontrés, la valeur axiologique de la base et la nature de la collocation. Les données analysées sont donc trop hétéroclites pour statuer sur un mécanisme stable, empêchant ainsi la formulation d'hypothèses interprétatives. Toutefois, deux éléments sont notoires, l'un en lien avec l'absence totale de mécanisme de transgression, ce qui pourrait représenter une manifestation possible de ce que nous avons appelé un « contrat de transparence », l'autre concernant la régularité des mécanismes de renforcement positif-négatif et d'activation également positif-négatif, appliqués aux collocations bivalentes, tout type confondu. Sans surprise, ces résultats nous ont permis de valider l'hypothèse posée comme générale, relative au fondement de la SPA concernant l'inscription de la subjectivité discursive dans les possibles argumentatifs des unités lexicales qui composent le discours scientifique spécialisé. Par répercussion, nous pouvons également valider l'hypothèse, qualifiée de triviale, d'une subjectivité réelle du discours scientifique spécialisé, étant entendu qu'y répondre par l'affirmative présente assez peu d'intérêt, ce qui importe étant de savoir à quel niveau cette subjectivité s'inscrit et à quels mécanismes elle obéit, ce à quoi nous sommes parvenus.

Conclusion et perspectives

Conclusion

Depuis son origine dans les années 1970, la linguistique de corpus mène une réflexion sur la pratique de corpus, en cherchant à faire de celui-ci un objet d'étude à part entière. Entre linguistique et informatique, la linguistique de corpus pose la question du rapport entretenu avec la recherche d'attestations en corpus textuel. L'analyse semi-automatique des données permet de traiter une grande quantité d'informations, offrant ainsi des perspectives de recherches jusque-là inexplorées. De fait, les observables produits par la pratique de corpus sont nouveaux, car attestés. En ce sens, ils posent la question du rapport entre empirie et théorie, tant en amont du processus de recherche dans la formulation des hypothèses qu'en aval dans la recherche de régularités et/ou de singularités. Cette approche empirique a permis de remplacer l'introspection fondée sur l'intuition du chercheur par une approche inductive, basée sur l'information rendue accessible par les textes composant le corpus. Or l'émergence de régularités textuelles est propre à constituer – à recréer – une norme linguistique, sociale, culturelle... Aussi, l'observation des données textuelles attestées abolit la dualité trop stricte entre langue et parole, héritée de Saussure, au profit d'analyses pluridimensionnelles, voire purement discursives. Dans ce contexte, l'intérêt de la sémantique pour la linguistique de corpus se justifie par l'émergence de problématiques inattendues, orientées vers la confrontation de modèles théoriques visant la mise à jour du « système » de la langue avec la réalité de la variation langagière. Pour opposées que soient ces deux approches, leur combinaison réinvente le *rapport* entre signification et sens par l'intermédiaire du discours et de sa norme, rendus accessibles par le corpus textuel.

De ce point de vue, la sémantique des possibles argumentatifs – SPA – présente un intérêt certain, puisque la description de la signification lexicale proposée intègre simultanément des informations sémantiques internes à la langue et des informations discursives externes, donc contextuelles. À l'interface entre la sémantique logique et structuraliste d'un côté et la sémantique historico-psychologique et cognitive de l'autre, la SPA réconcilie ces deux traditions linguistiques en adhérant à un processus de stabilisation du monde dans et par la langue. Ce processus est simultanément subjectif et intersubjectif, au sens où il est inévitablement variable dans l'espace individuel et social. Au niveau sémantique, la SPA postule que la langue générale et *a fortiori* la signification lexicale sont dénotatives, c'est-à-dire qu'elles résultent d'un besoin de dénomination rendu possible par une capacité de fixation référentielle ou d'habitudes associatives entre l'objet X et le signe linguistique x. Ce faisant, la SPA attribue à la signification lexicale l'existence d'un noyau contenant des informations stables sur lesquelles peuvent se construire les enchaînements argumentatifs, décrits en termes de blocs d'argumentations internes, les stéréotypes, et externes, les possibles argumentatifs – PA. À la différence de leur représentation respective indépendante en sémantique argumentative, les blocs d'argumentations internes et externes sont articulés en SPA. Cette description tridimensionnelle de la signification lexicale permet de mettre à jour des mécanismes explicatifs quant au cinétisme de la signification lexicale et par là de rendre

compte du pouvoir qu'a le discours de véhiculer des systèmes de valeur, de les transformer, mais aussi de les produire. Au niveau discursif, la SPA fonctionne en interaction nécessaire avec l'analyse linguistique du discours – ALD –, dont les préoccupations portent sur les mécanismes agissant sur les valeurs portées par les unités lexicales dans le discours. Ces mécanismes sont distincts selon l'axiologisation du mot analysé. Pour les monovalents positifs ou négatifs, il s'agit des six mécanismes définis par le modèle : le renforcement, l'activation, la neutralisation, l'affaiblissement, l'interversion et la transgression. Tandis que les mécanismes pour les bivalents se réduisent au renforcement et à l'activation dans leur version positive ou négative.

Nous avons vu, par ailleurs, que l'approche contextualiste développée par Firth désigne par *collocations* des phénomènes linguistiques de cooccurrence qui engendrent des modifications de sens constatées sur les mots, par leur relation de collocation en contexte. L'analyse des collocations se fonde donc sur l'observation des relations sémantico-discursives à la surface des textes. Aussi, cette approche revêt de notre point de vue un caractère mixte, semblable à celui proposé par le modèle SPA, puisque l'analyse des collocations se fonde également sur l'observation empirique des relations sémantiques – dimension sémantique –, créées par le contexte – dimension discursive. Or les problématiques de recherche concernant les collocations, qu'elles soient appliquées à divers domaines ou théoriques, visent toutes soit la création d'une modélisation de l'ensemble des phénomènes collocationnels, soit l'émergence des mécanismes inhérents à leur formation. Aussi, les préceptes théoriques de ces modèles reposant tous deux sur une interaction nécessaire entre le sémantique et le discursif, de ce rapprochement est venue l'idée de décrire argumentativement la motivation des entités mobilisées dans certains phénomènes collocationnels, en lien avec le pouvoir explicatif et prédictible potentiel de la SPA. L'approche contextualiste étudie les collocations dites textuelles, eu égard au phénomène de cohésion textuelle, inhérent à toute unité sémantique formée par le texte, dont ces collocations rendent compte et auquel elles participent. Pour les mêmes raisons, l'hypothèse d'une motivation argumentative des entités collocationnelles a désigné les collocations textuelles comme objet d'étude. Dans la lignée contextualiste, notre acception des phénomènes collocationnels correspond à une définition qualifiée de « large », au sens où elle repose essentiellement sur des critères statistiques que nous avons instanciés qualitativement et sur une typologie sémantique que nous empruntons à Cowie et qui distingue entre les collocations libres et les collocations restreintes. De façon traditionnelle, nous avons analysé uniquement des collocations lexicales, sans que le critère de bonne formation syntaxique n'intervienne dans la sélection des données.

L'origine des hypothèses formulées dans ce travail tient à une intuition issue du rapprochement des définitions de collocation textuelle et de possible argumentatif. Nous pensons en effet qu'une collocation textuelle correspond potentiellement à la concrétisation d'un PA dans le discours, ce qui revient à formuler une hypothèse sémantico-discursive sur la nature de l'association qui unit le collocatif à la base. Cette intuition nous a permis de poser une triple hypothèse permettant de statuer sur la validité d'une motivation argumentative sous-jacente à l'association collocationnelle, en postulant l'existence de collocations argumentatives, de collocations argumentativement motivées et de collocations arbitraires. De par le lien argumentatif posé, par essence évaluatif, nous pensons également que les collocations argumentatives et les collocations argumentativement motivées sont intrinsèquement axiologisées, orientées vers l'un ou l'autre des pôles axiologiques positifs ou négatifs. En l'absence de relation argumentative, la collocation arbitraire ne devrait pas être

axiologisée, ce que nous avons discuté. En effet, sans remettre en cause la dimension évaluative de la langue générale, nous pensons que, bien que neutralisée sémantiquement, la base conserve son potentiel axiologique. Ensuite, conformément aux postulats de la SPA, une seconde hypothèse allie les dimensions sémantico-discursives et pragmatico-discursives mobilisées par les six mécanismes d'axiologisation du sens au niveau discursif général cotextuel et/ou contextuel, dont nous pensons qu'ils sont transposables à la description des relations qui régissent l'association entre la base et le collocatif, relations que nous avons qualifiées de macro-sémantiques.

D'un point de vue pratique, nous nous sommes dotés d'un corpus électronique spécialisé, le corpus TAL(N), pour tester ces hypothèses. Le choix d'un tel corpus se justifie d'abord pour éprouver le fondement de la SPA en lien avec de la subjectivité discursive inscrite dans les PA des mots qui composent tout discours, postulat général posé en seconde « hypothèse » ; cette perspective implique d'ailleurs de rappeler préalablement l'objectivisme supposé du discours scientifique spécialisé, précepte trivial posé en première « hypothèse ». Il se justifie ensuite au niveau de notre objet d'étude même, puisque les collocations assurent un rôle dans la cohésion textuelle et sont susceptibles de pouvoir rendre compte des régularités et/ou singularités du corpus textuel constitué. Il se justifie enfin au niveau ALD, pour mettre à jour les mécanismes discursifs propres au discours TAL(N), dont nous pensons intuitivement qu'il « ne joue pas » à l'intervention des valeurs axiologiques, eu égard à un contrat de transparence et de non-ambiguïté dans la transmission du savoir.

Le corpus TAL(N), représentatif de la communauté du traitement automatique du langage naturel, a été constitué *via* les six critères proposés par la théorie des communautés de discours, dont les atouts reposent sur une délimitation de la communauté visée, de l'intérieur, sans recourir à des critères externes plus subjectifs liés au registre. En outre, cette démarche désigne objectivement plusieurs sources potentielles de données textuelles comme étant représentatives de la communauté, parmi lesquelles les conférences associées, les journées d'études, etc. En nous appuyant sur les critères traditionnels de constitution de corpus, cumulés à la fiabilité des critères induits par le modèle des communautés de discours, nous avons pu minimiser les effets subjectifs des critères externes de sélection des textes restant à notre charge, en ne retenant que les articles de la revue TAL et de la conférence TALN.

D'un point de vue méthodologique, nous avons mis en place un protocole de validation des hypothèses qui suit trois étapes : le repérage des unités lexicales, le dépouillement des données et la construction de la signification lexicale. Le repérage des unités lexicales a d'abord fait l'objet d'une approche déductive ; les mots à analyser ont été choisis en fonction de la subjectivité supposée du discours TAL. Or la qualité et la quantité de données récoltées se sont révélées si pauvres que nous avons adopté en définitive une démarche inductive inverse. L'approche inductive suppose de sélectionner les données en fonction de leur présence significative dans le corpus et de leur pertinence au regard de l'hypothèse posée sur la motivation argumentative. Cette pertinence repose sur l'axiologisation positive ou négative des unités sélectionnées en qualité de base. Dès lors, nous avons choisi de construire un paradigme expérimental équilibré, composé de deux monovalents positifs – *améliorer-amélioration(s)* et *performance(s)* –, de deux monovalents négatifs – *erreur(s)* et *problème(s)* – et de deux bivalents – *évaluer-évaluation(s)* et *résultat(s)*. Pour le dépouillement des données, nous avons rassemblé quatre critères : un critère statistique, un critère de pertinence, un critère lexical et un critère terminologique ; la fiabilité du dépouillement tenant à l'application simultanée de ces quatre critères à un même mot. Concernant l'aspect

lexicographique, la construction de la signification lexicale a été effectuée à partir des définitions de chaque mot pris dans deux dictionnaires différents : *Le Grand Robert* et *Le Larousse*. Pour ce faire, nous avons établi une nomenclature des concepts et des termes associés dans ces deux dictionnaires. Cette démarche s'est révélée fructueuse, puisqu'elle a permis d'identifier trois concepts lexicographiques stables (l'entrée, la définition et les exemples) nécessaires à l'élaboration de la signification lexicale : les PA correspondent systématiquement aux exemples d'emplois proposés pour le mot auquel est consacré l'article. Enfin, les deux tests argumentatifs proposés, l'un de substitution – TSA –, l'autre d'enchaînement – TEA –, ont permis d'évaluer les données recueillies dans le corpus TAL(N).

L'application de ces différents principes méthodologiques valide la plupart de nos hypothèses de recherche. Concernant le paradigme d'hypothèses H3, conformément à nos attentes, les analyses sémantico-discursives menées sur les six bases retenues confortent l'hypothèse relative à l'existence de collocations argumentatives, quelle que soit la nature axiologique de la base : monovalente positive (**Ex** : *gain de performance*) vs négative (**Ex** : *diagnostic d'erreur*) ou bivalente (**Ex** : *évaluation comparative*). Dans ce cas, l'adjonction du collocatif à la base forme avec les stéréotypes proposés ou envisageables un bloc d'argumentation interne. Elles correspondent à la concrétisation d'un PA dans le discours. Ces collocations sont prévisibles, « calculables » sur la base des PA existants. Conformément aux postulats théoriques de la SPA-ALD, de par le lien argumentatif posé par la formation d'un bloc d'argumentation externe, les collocations argumentatives sont intrinsèquement axiologisées, c'est-à-dire orientées vers l'un ou l'autre des pôles axiologiques positif (**Ex** : *bonne performance*) ou négatif (**Ex** : *problème épineux*).

Les analyses sémantico-discursives effectuées ont également permis de valider la possibilité de collocations argumentativement motivées, les plus fréquentes (**Ex** : *amélioration sensible* ; *réduction des erreurs* ; *résultats de la recherche*). Pour toutes les collocations concernées, l'application du TSA a montré que l'adjonction du collocatif à la base forme un déploiement argumentatif possible. Toutes ces collocations entretiennent une relation argumentative avec la base, ce qui permet de décrire et d'expliquer le principe sous-jacent à l'association des deux entités formant la collocation. Le mécanisme de motivation argumentative revêt un caractère moins prédictible que celui mobilisé par les collocations argumentatives. Toutefois, l'émergence de principes associatifs récurrents en fonction de la monovalence positive-négative ou de la bivalence de la base est possible. Là aussi, de par le lien argumentatif déployé, ces collocations comportent une évaluation orientée vers l'un ou l'autre des pôles axiologiques, positif (**Ex** : *amélioration substantielle*) ou négatif (**Ex** : *risque d'erreur*).

Enfin, l'existence de collocations arbitraires est également vérifiée. En nombre plus restreint dans le corpus, ces collocations ne correspondent ni à la concrétisation d'un PA dans le discours ni à un cas de motivation argumentative. Dans la mesure où l'adjonction du collocatif à la base ne forme pas avec les stéréotypes proposés ou envisageables un bloc d'argumentation interne ou un déploiement argumentatif possible, ces collocations sont arbitraires (**Ex** : *amélioration de la méthode* ; *problème sémantique* ; *évaluation automatique*). Elles ne peuvent être sémantiquement prédictibles et n'ont de réalité que discursive. Nous avons pu constater que l'absence de lien argumentatif interne à la collocation ne l'empêchait nullement de conserver un potentiel axiologique, hérité de la base quand celle-ci est axiologiquement marquée, hérité du collocatif quand la base est bivalente.

Ce faisant, les analyses menées ont révélé un parallèle intéressant concernant la superposition possible de la typologie sémantique, collocations libres vs collocations restreintes adoptée

pour classifier les collocations textuelles et la typologie argumentative obtenue par la validation de notre hypothèse centrale. En effet, il est apparu que les collocations libres correspondent aux cas de motivations argumentatives décrits dans notre étude, de même que les collocations restreintes se rapprochent des collocations argumentatives. Quant aux collocations arbitraires, elles présentent la particularité d'être représentatives des pratiques liées au domaine TAL, instanciant ainsi pleinement leur fonction cohésive au sein du corpus représentatif de cette communauté de discours.

De même, l'hypothèse d'une transposition possible des mécanismes d'axiologisation du sens au niveau discursif général cotextuel et/ou contextuel au niveau macro-sémantique s'est révélée concluante. Les analyses menées ont fait apparaître des mécanismes récurrents, comparables à ceux dégagés par les études menées sur la motivation sémantique. Les collocations textuelles étant peu imaginées, les régularités argumentatives observées corroborent simultanément les mécanismes de « convergence de polarité » et de « renforcement » inhérent au principe de « congruence sémantique ». Quels que soient les mécanismes sémantico-discursifs ou pragmatico-discursifs mobilisés par les collocations argumentatives, motivées ou arbitraires, quatre mécanismes macro-sémantiques sont récurrents et suffisent à expliquer la majorité des données recensées ; il s'agit des mécanismes de renforcement, d'activation, de neutralisation (stabilisation) et d'interversion. Toutefois, nous avons pu observer quelques phénomènes collocationnels *a priori* réticents à toute classification. Dans ces cas, les nombreux paramètres intervenant dans l'analyse et leurs multiples combinaisons empêchent de statuer sur un mécanisme stable. Cependant, les éléments de réponse s'orientent sur la nature grammaticale du collocatif, son antéposition ou sa postposition par rapport à la base et l'aspect normatif ou transgressif du déploiement argumentatif sous-tendant l'association collocationnelle. D'un point de vue discursif général, les résultats des analyses sont moins probants. Nous pensions intuitivement trouver une majorité de potentiels activés, voire renforcés, en particulier pour les collocations arbitraires, mais pas ou peu de potentiels neutralisés ou affaiblis et encore moins de potentiels intervertis ou transgressés, eu égard à un contrat de transparence et de non-ambiguïté dans la transmission du savoir. Or nous avons été surpris par l'absence de régularité et, *a contrario*, par la multitude de combinaisons possibles entre les mécanismes discursifs rencontrés, la valeur axiologique de la base et la nature de la collocation. Les données analysées sont donc trop hétéroclites pour statuer sur un mécanisme stable, empêchant ainsi la formulation d'hypothèses interprétatives. Toutefois, deux éléments sont notoires, l'un en lien avec l'absence totale de mécanisme de transgression, ce qui pourrait représenter une manifestation possible de ce que nous avons appelé un « contrat de transparence », l'autre concernant la régularité des mécanismes de renforcement positif-négatif et d'activation également positif-négatif, appliqués aux collocations bivalentes, tout type confondu. Sans surprise, ces résultats nous ont permis de valider l'hypothèse posée comme générale, relative au fondement de la SPA concernant l'inscription de la subjectivité discursive dans les possibles argumentatifs des unités lexicales qui composent le discours scientifique spécialisé. Par répercussion, nous pouvons également valider l'hypothèse, qualifiée de triviale, d'une subjectivité réelle du discours scientifique spécialisé, étant entendu qu'y répondre par l'affirmative présente assez peu d'intérêt, ce qui importe étant de savoir à quel niveau cette subjectivité s'inscrit et à quels mécanismes elle obéit, ce à quoi nous sommes parvenus.

Dans la lignée des problématiques théoriques ayant à cœur d'étudier les mécanismes permettant de décrire – de générer –, voire de prédire, les phénomènes collocationnels, le modèle SPA nous a permis de valider une hypothèse sur la nature *argumentative* de la

relation qui unit la base au collocatif dans les collocations textuelles. Ce faisant, nous avons mis à jour un même principe descriptif-explicatif, voire prédictible, pour les collocations libres (dans les cas de motivations argumentatives) et les collocations restreintes (dans les cas de collocations argumentatives). En effet, les collocations libres ne semblent pas « réellement » libres, au sens où leur association est souvent contrainte par un principe argumentatif, principe possiblement sous-jacent à leur prédictibilité, autre que psychologique ; tandis que la prédictibilité des collocations restreintes repose sur un nouveau mécanisme de génération, autre que syntaxique et/ou sémantique.

Perspectives

Au niveau théorique, quoique traditionnellement le phénomène collocationnel soit décrit par son caractère binaire [Tutin et Grossmann 2002], la perspective d'une analyse portant sur des collocations comprenant plus de deux unités lexicales nous semble intéressante. Du point de vue argumentatif, les collocations argumentatives et les collocations argumentativement motivées, de par le bloc ou le déploiement qu'elles contiennent, sont propres à intégrer des phénomènes d'argumentations séquentielles ou par visée, les enchaînements discursifs opérant autant à partir des énoncés et des syntagmes qu'à partir de mots [Carel et Ducrot 1999]. Par ailleurs, au niveau macro-sémantique, les informations recueillies permettraient vraisemblablement d'obtenir des informations supplémentaires sur le rôle joué par la base à l'égard des mécanismes sémantico-discursifs et pragmatico-discursifs d'axiologisation du sens. Enfin, du point de vue théorique également, une autre perspective d'analyse consiste à interroger la circularité du processus argumentatif, inadapté à la description des collocations. En effet, la relation d'argumentation est définie par la relation entre deux entités, un argument et une conclusion, posée comme naturelle. Aussi peut-on dire *Processus DC Résultat*, mais également *Résultat DC processus*. Pour autant, transposée aux phénomènes collocationnels, la réversibilité n'est pas équivalente. Parler d'un *problème de réponse*, ce n'est pas parler d'une *réponse au problème*. Or, dans l'état actuel du protocole de validation, le test de substitution argumentatif ne permet pas de distinguer entre ces deux blocs d'argumentations, en véhiculant leur sens respectif. Toutefois, nous n'avons pas été confrontés directement à ce problème dans les analyses menées, étant donné que les « doublons » rencontrés concernaient des collocations non pas argumentatives mais arbitraires, comme *pourcentage d'erreur* et *erreur de pourcentages*.

Au niveau empirique, nous avons établi une nomenclature très simple des concepts et des termes associés dans les dictionnaires *Le Grand Robert* et *Le Larousse*. Cette nomenclature comporte l'article, l'entrée, la définition et l'exemple. Or, à l'usage, il s'est révélé que, pour la construction de la signification lexicale de chacun des six mots étudiés, les paramètres lexicographiques mobilisés étaient identiques. C'est-à-dire que, pour le noyau, par exemple, composé de prédicats abstraits, nous ne retenons pour sa construction que les définitions, parmi lesquelles celles qui sont communes aux différentes entrées. En outre, toutes les définitions spécialisées à un domaine de la connaissance ou à une communauté professionnelle sont exclues. Aussi, la stabilité des informations nécessaires à l'élaboration de la signification lexicale nous paraît propice à la mise en place d'un protocole de construction automatique de la signification lexicale, à partir de deux bases lexicographiques.

À l'interface entre théorie et empirie, nous proposons une dernière perspective de recherche. Nous avons vu que la SPA est un modèle de description de la signification lexicale mixte indexé simultanément sur des informations sémantiques et discursives. C'est pourquoi cette approche permet, au niveau sémantique, de mettre à jour des mécanismes explicatifs quant au cinétisme de la signification lexicale et par là de rendre compte du pouvoir qu'a le discours de véhiculer des systèmes de valeurs, de les transformer, mais aussi de les produire. Par voie de conséquence, l'hypothèse posée quant aux mécanismes discursifs d'axiologisation du sens ayant été peu concluante, nous aimerions dans un premier temps partir du corpus pour reconstruire les significations lexicales des unités sélectionnées dans le discours scientifique spécialisé, à partir des enchaînements discursifs présents en corpus [Galatanu 2005g, à paraître]. Ce faisant, nous aimerions pouvoir formuler des hypothèses interprétatives sur la pratique TAL qui sous-tend le discours étudié, ce qui est précisément rendu possible par la linguistique de corpus : « Le “ chaînon manquant ” entre la langue et la parole est constitué par l'espace des normes (*cf.* Coseriu, 1969). Or seule la linguistique de corpus peut offrir les moyens théoriques et techniques d'étudier l'espace des normes et de transformer en dualité l'antinomie entre compétence et performance. » [Rastier 2005, p. 34]

Nous avons vu en introduction que tout le monde concevait désormais le corpus comme un *observable*, sans pour autant s'accorder sur une approche commune : la *linguistique de corpus* penche du côté de la description interprétative, de l'induction et de la parole, contre la *linguistique sur corpus*, qui penche du côté de l'illustration par l'exemple, de la déduction et de la langue [Mayaffre 2005, p. 12]. C'est la distinction que Williams établit entre les études *corpus-based* et les études *corpus-driven* [Williams 2005, p. 13].

Conformément à cette distinction, le travail achevé relève en premier lieu d'une linguistique *sur corpus*, tandis que la recherche à venir sur la reconstruction de la signification lexicale des entités à partir des déploiements discursifs observés s'oriente vers une linguistique *de corpus*. Selon les stades de la recherche en cours, tantôt la théorie investit l'empirie, tantôt c'est l'inverse, dans un processus de circularité progressive et cumulative. C'est pourquoi « il faut donc passer du principe de plaisir théorique au principe de réalité philologique. Un nouveau rapport à l'empirique change non seulement l'étendue mais la nature des faits et rend nécessaire l'innovation théorique. » [Rastier 2005, p. 41]. C'est précisément ce que nous entendons poursuivre dans nos travaux futurs.

Bibliographie

- [Aarts 1990] J. Aarts, Corpus Linguistics: an appraisal, in *Computers in Literary and Linguistics Research*, J. Hamesse and A. Zampolli (eds.), Paris, Champion Slatkine, pp.13-28.
- [Adam 1992] J.-M. Adam, *Les textes : types et prototypes*, Paris, Nathan.
- [Alonso Ramos 2003] M. Alonso Ramos, La nature des collocatifs : leur statut en tant qu'unités lexicales, dans *Les collocations : analyse et traitement*, F. Grossmann et A. Tutin (éds.), Travaux et recherches en linguistique appliquée, Amsterdam, de Werelt, pp. 45-60.
- [Anscombre 1989] J.-C. Anscombre, Théorie de l'argumentation, topoï et structuration discursive, *Revue Québécoise de linguistique*, Vol. 18(1), pp. 13-56.
- [Anscombre 1995a] J.-C. Anscombre, De l'argumentation dans la langue à la théorie des topoï, dans J.-C. Anscombre (éd.), *Théorie des topoï*, Paris, Kimé, pp. 11-47
- [Anscombre 1995b] J.-C. Anscombre, La nature des topoï, dans J.-C. Anscombre (éd.), *Théorie des topoï*, Paris, Kimé, pp. 49-84.
- [Anscombre 1995c] J.-C. Anscombre, *Théorie des topoï*, J.-C. Anscombre (éd.), Paris, Kimé.
- [Anscombre et Ducrot 1983], J.-C. Anscombre et O. Ducrot, *L'argumentation dans la langue*, Bruxelles, Pierre Mardaga.
- [Aston and Burnard 1998] G. Aston and L. Burnard, *The BNC Handbook: Exploring the British National Corpus with SARA*, Edinburgh, Edinburgh University Press.
- [Atkins and al. 1992] B. Atkins, J. Clear, N. Ostler, Corpus Design Criteria: Literary and linguistic Computing, *Journal of the Association for Literary and linguistic Computing*, Vol. 7(1), pp. 1-16.
- [Bachimont 2005] B. Bachimont, Corpus et connaissances : de l'extraction linguistique à la modélisation conceptuelle, dans A. Condamines (éd.), *Sémantique et corpus*, Paris, Hermès, Lavoisier, pp. 319-344.
- [Bachschmidt 1997] P. Bachschmidt, Procédure de constitution d'un corpus attesté d'articles de recherche scientifique, *Revue ASp*, Vol. 15-18, p.133-138.
- [Bahns 1993] J. Bahns, Lexical collocations: a contrastive view, *ELT Journal*, Vol. 47(1), pp. 56-63.
- [Bally 1909] C. Bally, *Traité de stylistique française*, Genève, Paris, Georg et Klincksieck.
- [Bally 1932] C. Bally, *Linguistique générale et linguistique française*, Berne, Francke.
- [Baranov and Dobrovolskii 1996] A.N. Baranov and D.O. Dobrovolskii, Cognitive modeling of actual meaning in the field of phraseology, *Journal of Pragmatics*, Vol. 25(3), pp. 409-429.

- [Beauchesne 2001] J. Beauchesne, *Dictionnaires des cooccurrences*, Montréal – Toronto, Guérin.
- [Beaune 1988] J.-C. Beaune, La vulgarisation scientifique. L'ombre des techniques, dans D. Jacobi, B. Schiele (éds.), *Vulgariser la science*, Seyssel, Éditions Champ Vallon (Milieux), pp. 47-81.
- [Benson 1989] M. Benson, The Structure of the Collocational Dictionary, *International Journal of Lexicography*, 2-1, pp. 1-14.
- [Benson et al. 1986] M. Benson, E. Benson and R. Ilson, *The BBI Dictionary of English Word Combinations*, Amsterdam, John Benjamins (2nd edition, 1997).
- [Berry-Roghe 1973] G.L.M. Berry-Roghe, The computation of Collocations and their Relevance in Lexical Studies, in A.J. Aitken, R. Bailey and N. Hamilton-Smith (eds.), *The Computer and Literary Studies*, Edinburgh, Edinburgh University Press .
- [Beyssade 1998] C. Beyssade, *Sens et savoirs : des communautés épistémiques dans le discours*. Presses Universitaires de Rennes.
- [Bathia 1993] V. K. Bhatia, *Analysing Genre: Language Use in Professional Settings*, London, Longman.
- [Benson 1985] M. Benson, Collocations and Idioms, in R. Ilson (ed.), *Dictionaries, Lexicography and Language Learning*, Oxford, Pergamon, The British Council, pp. 61-68.
- [Biber 1988] D. Biber, *Variation across speech and writing*, Cambridge, Cambridge University Press.
- [Biber 1989] D. Biber, A typology of English texts, *Linguistics*, N°27, pp. 3-43.
- [Biber 1993] D. Biber, Representativeness in Corpus Design; Literary and linguistic Computing, *Journal of the Association for Literary and linguistic Computing*, Vol. 8(4), pp. 243-257.
- [Biber 1995] D. Biber, *Dimensions of register variation: a cross-linguistic comparison*. Cambridge, Cambridge University Press.
- [Biber and Finegan 1986] D. Biber and E. Finegan, An Initial Typology of English Text Types, in *Corpus Linguistics II : New Studies in the Analysis and Exploitation of Computer Corpora*, J. Aarts and W. Meijs (dir.), Vol. 47, Amsterdam, Rodopi, pp. 19-46.
- [Biber et al. 1994] D. Biber, E. Finegan and D. Atkinson, Archer and its challenges: compiling and exploring a representative corpus of historical English registers, in *Creating and using English language corpora*, U. Fries, G. Tottie and P. Schneider (eds.), Amsterdam, Rodopi, pp. 1-14.
- [Biber et al. 1998] D. Biber, S. Conrad and R. Reppen, *Corpus Linguistics: Investigating Language Structure and Use*; Cambridge, Cambridge University Press.
- [Binon et al. 2000] J. Binon, S. Verlinde, J. Van Dyck et A. Bertels, *Dictionnaire d'apprentissage du français des affaires*, Paris, Didier.

- [Burnard et Sperberg 1996] L. Burnard et M.-C. Sperberg, La TEI simplifiée : une introduction au codage des textes électroniques en vue de leur échange, *Cahiers de GUTenberg*, N° 24, pp. 23-151.
- [Caignon 2001] P. Caignon, *Vocabulaire et cooccurents de la publicité*, Montréal, Linguattech.
- [Campione et al. 2005] E. Campione, J. Véronis and J. Deulofeu, The French corpus, in E. Cresti and M. Moneglia (eds.), *C-ORAL-ROM, Integrated Reference Corpora for Spoken Romance Languages*, Amsterdam, John Benjamins, pp. 111-133.
- [Carel 1995] M. Carel, Trop: argumentation interne, argumentation externe et positivité, sous la direction de J-C. Anscombre, *Théorie des topoï*, Paris, Kimé, pp.177-206.
- [Carel et Ducrot 1999] M. Carel et O. Ducrot, Le problème du paradoxe dans une sémantique argumentative, dans O. Galatanu et J.-M. Gouvard (éds.), *Langue Française 123: La sémantique du stéréotype*, pp. 6-26.
- [Chomsky 1957] N. Chomsky, *Syntactic Structures*, La Hague, Mouton.
- [Chomsky 1965] N. Chomsky, *Aspect of the Theory of Syntax*, MIT Press, Cambridge, Massachusetts.
- [Church and Hanks 1990] K.W. Church and P. Hanks, Word Association Norms, Mutual Information, and Lexicography, *Computational Linguistics*, 16(1), pp. 22-29.
- [Clear 1993] J. Clear, From Firth principles: Computational tools for the study of collocation, in M. Baker, G. Francis, and E. Tognini-Bonelli (eds.), *Text and technology: In honour of John Sinclair*, Amsterdam, John Benjamins.
- [Clear 1994] J. Clear, I can't see the sense in a large corpus, *COMPLEX '94 - Proceedings of the 3rd International Conference on Computational Lexicography*, Hungary, Budapest, pp. 33-48.
- [Cohen 1986] B. Cohen, Lexique de cooccurents – *Bourse et conjuncture économique*, Montréal, Linguattech.
- [Condamines 2005] A. Condamines, Sémantique et corpus, quelles rencontres possibles ?, dans A. Condamines (éd.), *Sémantique et corpus*, Paris, Hermès, Lavoisier, pp. 15-36.
- [Condamines et Rebeyrolle 1996] A. Condamines et J. Rebeyrolle, Point de vue en langue spécialisée, *META*, Vol. 42(1) pp. 174-184.
- [Corbin 1997] D. Corbin, Locutions, composés, unités polylexématiques : lexicalisation et mode de construction, dans M. Martins-Baltar (éd.), *La locution, entre langue et usages*, Fontenay-St-Cloud, ENS Editions, pp. 55-102.
- [Coseriu 1969] E. Coseriu, Sistema, norma, et 'parola', *Studi linguistici in onore Vittorio Pisani*, Brescia, Paideia, pp. 235-253.
- [Coseriu 2000] E. Coseriu, *L'homme et son langage*, Louvain, Sterling, Peeters.
- [Cowie 1978] A. Cowie, The place of illustrative material and collocations in the design of a learner's dictionary, in P. Strevens (ed.), *In Honour of A.S. Hornby*, Oxford, Oxford University Press.

- [Cowie 1998] A. Cowie, *Phraseology: Theory, Analysis and Applications*, A.P. Cowie (ed.) Oxford, Clarendon Press.
- [Cowie et al. 1983] A. Cowie, R. Mackin and I. R. McCaig, *Oxford Dictionary of Current Idiomatic English, Vol. 2 - English Idioms*, Oxford, Oxford University Press.
- [Cruse 1986] D. A. Cruse, *Lexical Semantics*, Cambridge, Cambridge University Press.
- [Daille 1996a] B. Daille, ACABIT : une maquette d'aide à la construction automatique de banques terminologiques monolingues ou bilingues, dans A. Clas, P. Thoiron et H. Béjoint (éds), *Lexicomatique et Dictionnairiques*, FMA, Beyrouth, pp. 123-136.
- [Daille 1996b] B. Daille, Study and Implementation of Combined Techniques for Automatic Extraction of Terminology, in P. Resnik and J. Klavans (eds), *The Balancing Act: Combining Symbolic and Statistical Approaches to Language*, MIT Press, pp. 49-66.
- [Daille et Romary 2001] B. Daille et L. Romary (éds.), *Linguistique de corpus*, Traitement Automatique des Langues (TAL), Vol. 42(2), Hermès Sciences Publications.
- [Daille and Williams 2001] B. Daille and G. Williams (eds.), *COLLOCATION: Computational Extraction, Analysis and Exploitation*, Workshop in ACL'01, France, Toulouse.
- [Dubreil et Daille 2005a] E. Dubreil et B. Daille, Analyse sémantico-discursive des collocations lexicales en corpus spécialisé : la base « connaissance-s », « Mots, Termes et Contextes », *Actes des 7^{èmes} journées scientifiques LTT*, 6-10 septembre 2005, Bruxelles, (à paraître).
- [Dubreil 2005b] E. Dubreil, Principe de composition sémantico-discursif des collocations lexicales : de l'arbitraire au motivé ?, *Actes des 4^{èmes} Journées de la Linguistique de Corpus (JLC)*, 15-17 septembre 2005, Lorient, Université de Bretagne Sud, (à paraître).
- [Ducrot 1972] O. Ducrot, *Dire et ne pas dire*, Paris, Hermann.
- [Ducrot 1973] O. Ducrot, *La Preuve et le Dire*, Repère, Paris, Mame.
- [Ducrot 1980] O. Ducrot, *Les mots du discours*, Paris, Minuit.
- [Ducrot 1984] O. Ducrot, *Le dire et le dit*, Paris, Minuit.
- [Ducrot 1993] O. Ducrot, Sémantique des langues naturelles, *Séminaire 1993-1994 de l'EHESS*.
- [Ducrot 1995a] O. Ducrot, Topoi et formes topiques, dans J.-C. Anscombe (éd.), *Théorie des topoi*, Paris, Kimé, pp. 85-100.
- [Ducrot 1995b] O. Ducrot, Les modificateurs déréalisants, *Journal of Pragmatics*, Vol. 24, pp. 145-165.
- [Ducrot et Carel 1999] O. Ducrot et M. Carel, Les propriétés linguistiques du paradoxe : paradoxe et négation, dans O. Galatanu et J.-M. Gouvard (éds.), *Langue Française 123: La sémantique du stéréotype*, pp. 27-40.
- [Ducrot et Schaeffer 1995] O. Ducrot et J.-M. Schaeffer, *Nouveau dictionnaire encyclopédique des sciences du langage*, Paris, Seuil.

- [Engwall 1994] G. Engwall, Not chance but choice : criteria in corpus creation, in *Computational approaches to the lexicon*, B. Atkins and A. Zampolli (eds.), Oxford, Oxford University Press.
- [Eufe et al. 2004] R. Eufe, L. Fesenmeier, C. Ozouf et M. Senechal, Le dictionnaire électronique des synonymes du CIRSCO, J.-L. Manguin et J. François (coord.), *Cahier du CIRSCO*, N° 17.
- [Eurodicautom] <http://europa.eu.int/eurodicautom/Controller>
- [Firth 1957] J.R. Firth, Modes and Meaning, *Papers in linguistics* 1934-1951, Oxford, Oxford University Press.
- [Fontenelle 1997] T. Fontenelle, Turning a Bilingual Dictionary into a Lexical-Semantic Database, *Lexicographica*, Series Maior 79, Tübingen, Niemeyer.
- [Fontenelle 2003] T. Fontenelle, Collocations et traitement automatique du langage naturel, dans *Les collocations : analyse et traitement*, F. Grossmann et A. Tutin (éds.), Travaux et recherches en linguistique appliquée, Amsterdam, de Werelt, pp. 75-87.
- [Fourour 2001] N. Fourour, Identification et catégorisation automatique des anthroponymes du Français, *RECITAL 2001*, Tours, 2-5 juillet, pp. 441-451.
- [Galatanu 1999a] O. Galatanu, Le phénomène sémantico-discursif de déconstruction – reconstruction des topoï dans une sémantique argumentative intégrée, dans O. Galatanu et J.-M. Gouvard (éds.), *Langue Française 123: La sémantique du stéréotype*, pp. 41-51.
- [Galatanu 1999b] O. Galatanu, Argumentation et analyse du discours, dans Y. Gambier et E. Suomela-Salmi (éds.), *Jalons*, 2, Turku, Université de Turku, pp. 41-54.
- [Galatanu 2000a] O. Galatanu, Signification, sens et construction discursive de soi et du monde, dans J.M. Barbier, et O. Galatanu, (éds.), *Signification, sens, formation*, Paris, Presses Universitaires de France, pp. 25-44.
- [Galatanu 2000b] O. Galatanu, Langue, Discours et systèmes de valeurs, dans E. Suomela-Salmi (éd.), *Curiosités linguistiques*, Université de Turku, pp. 80-102.
- [Galatanu 2000c] O. Galatanu, La reconstruction du système de valeurs convoquées et évoquées dans le discours médiatique, dans *Actes du XXIIe congrès international de Linguistique et Philologie romanes*, Bruxelles, 23-29 juillet 1998, Vol. VII. Max Niemeyer Verlag, pp. 251-258.
- [Galatanu 2002a] O. Galatanu, Le mécanisme sémantico-discursif de l'écriture différée : le cas de Baudelaire, dans *Écritures différées*, Clermont-Ferrand, Presses Universitaires Blaise Pascal, pp. 175 -191.
- [Galatanu 2002b] O. Galatanu, La dimension axiologique de l'argumentation, dans M. Carel (éd.), *Hommage à Oswald Ducrot*, Paris, Kimé, pp. 93-107.
- [Galatanu 2002c] O. Galatanu, Le concept de modalité : les valeurs dans la langue et dans le discours, dans *Les valeurs*, Séminaire Le lien social, pp. 17-32.
- [Galatanu 2003a] O. Galatanu, La construction discursive des valeurs, dans J.-M. Barbier (éd.), *Valeurs et activités professionnelles*, Paris, L'Harmattan, pp. 87 -115.

[Galatanu 2003b] O. Galatanu, La sémantique des valeurs dans la prière française, dans J.-C. Aroui (éd.), *Le sens et la mesure. De la pragmatique à la métrique*. Hommages à Benoît de Cornulier, Paris, Champion Editeur, pp. 69-88.

[Galatanu 2004] O. Galatanu, La sémantique argumentative intégrée et ses enjeux pour l'analyse linguistique du discours, dans *Les actes du Congrès International d'Etudes Françaises : La Rioja, Croisée des Chemins* ; Vol. 2, M.J. Salinero Cascante et Ignacio Inarrea Las Veras (éds.), pp. 213-225.

[Galatanu 2005a] O. Galatanu, La sémantique des modalités et ses enjeux théoriques et épistémologiques dans l'analyse des textes, dans J.M. Gouvard (éd.), *De la langue au style*, Paris, Presses Universitaires de Lyon, pp. 157-170.

[Galatanu 2005b] O. Galatanu, Sémantique et élaboration discursive des identités, L'Europe de la connaissance dans le discours académique, dans les *Actes du colloque « Cross-cultural and cross-linguistic perspectives on Academic Discourse »* Mai 20-22, 2005, Université de Turku, Finlande, (à paraître).

[Galatanu 2005c] O. Galatanu, Pour une sémantique argumentative dans l'étude de la « proximité – distance » des systèmes lexicaux des langues romanes, dans les *Actes du colloque « Langues proches » (Langues collatérales 2)*, Université de Limerick – Université d'Amiens, Limerick, Irlande, 16-18 juin 2005, (à paraître).

[Galatanu 2005d] O. Galatanu, La « stéréophagie », un phénomène discursif de déconstruction – reconstruction de la signification lexicale », dans les *Actes du Colloque international « Représentation du sens linguistique III – RSL III »*, Bruxelles, 3-5 novembre 2005, (à paraître).

[Galatanu 2005e] O. Galatanu, Sémantique des possibles argumentatifs et axiologisation discursive, dans D. Bouchard, J. Evrard (éds.), *Représentation du Sens linguistique II*, Paris, Editions Duculot, Louvain-la-Neuve, (à paraître).

[Galatanu 2005f] O. Galatanu, Sémantique des possibles argumentatifs et dénomination, dans M. Riegel, C. Schnedecker, P. Swiggers et I. Tamba (éds.), *Au carrefour du sens*. Hommages offerts à Georges Kleiber, Louvain, Peeters, (à paraître).

[Galatanu 2005g] O. Galatanu, Le cinétisme de la signification lexicale, dans J.-M. Barbier, M. Durand (éds.), *Sujets, activité, environnement*, Paris, Presses Universitaires de France, (à paraître).

[Galatanu 2006] O. Galatanu, La dimension axiologique de la dénomination, dans M. Riegel, C. Schnedecker, P. Swiggers et I. Tamba (éds.), *Aux carrefours du sens*. Hommages offerts à Georges Kleiber, Louvain, Peeters, (à paraître).

[Gardies 1979] J.-L. Gardies, *Essai sur la logique des modalités*, Paris, Presses Universitaires de France.

[Gardies 1983] J.-L. Gardies, Tentative d'une définition de la modalité, dans J. David et G. Kleiber (éds.), *La notion sémantico-logique de modalité*, Université de Metz, pp. 13-24.

[GDT] Grand Dictionnaire Terminologique :

http://www.granddictionnaire.com/btml/fra/r_motclef/index1024_1.asp

- [Geeraerts 1991] D. Geeraerts, Grammaire cognitive et sémantique lexicale, dans *Sémantique cognitive, volume 58 of Communications*, Seuil, pp. 17-50.
- [Gonzalez Rey 2002] I. Gonzalez Rey, *La phraséologie du français*, Toulouse, Presses Universitaires du Mirail.
- [Goossens 1995] M. Goossens, Introduction pratique à SGML, *Cahiers de GUTenberg*, N° 19, pp. 27-58.
- [Grandchamp 1996] J.-M. Grandchamp, *L'argumentation dans le traitement automatique de la langue*, Thèse de doctorat de l'université Paris XI – Orsay, soutenue le 16/01/96.
- [Granger 1998] S. Granger, Prefabricated patterns in advanced EFL writing: collocations and formulae. In AP Cowie (ed.) *Phraseology: Theory, Analysis, and Applications*. Oxford, Clarendon Press, pp. 145-160.
- [Gross 1981] G. Gross, Lexicographie et Grammaire, *Cahiers de lexicologie*, Vol. 39(2), pp. 35-46.
- [Gross 1986] G. Gross, Les constructions nominales et l'élaboration d'un lexique-grammaire, *Langue Française*, N° 69, Paris, Larousse, pp. 5-27.
- [Gross 1991] G. Gross, Verbes supports et déterminant zéro, *Langages*, N° 102, Larousse, pp. 36-51.
- [Gross 1993] G. Gross, Trois applications de la notion de verbe support, *L'information grammaticale*, N° 59, Paris, pp. 16-23.
- [Gross 1994] G. Gross, Classes d'objets et description des verbes, *Langages*, N° 115, Larousse, Paris, pp. 15-31.
- [Gross 1995] G. Gross, Une sémantique nouvelle pour la traduction automatique : les classes d'objets, *La Tribune des Industries de la Langue et de l'Information électronique*, N° 17-19, Paris, pp. 16-19.
- [Gross 1996] G. Gross, *Les expressions figées en français. Noms composés et autres locutions*, Paris, Ophrys.
- [Grossmann et Tutin 2003] F. Grossmann et A. Tutin, Quelques pistes pour le traitement des collocations, *Les collocations : analyse et traitement*, F. Grossmann et A. Tutin (éds.), Travaux et recherches en linguistique appliquée, Amsterdam, de Werelt, pp. 5-21.
- [Grossmann et Tutin 2005] F. Grossmann et A. Tutin, Joie profonde, affreuse tristesse, parfait bonheur : sur la prédicativité des adjectifs intensifiant certains noms d'émotion. *Cahiers de lexicologie*, N° 86, 2005-1, pp. 1-18.
- [Grossmann and Tutin à paraître] F. Grossmann and A. Tutin, Motivation of Lexical Associations in Collocations : the Case of Intensifiers denoting 'Joy', dans *Festschrift In Honour of Igor Mel'cuk*, L. Wanner (ed.), Amsterdam, John Benjamins, (à paraître).
- [Habert 2000] B. Habert, Des corpus représentatifs: de quoi, pour quoi, comment ? dans M. Bilger (éd.), *Linguistique sur corpus. Etudes et réflexions*. Perpignan, Presses Universitaires de Perpignan, pp. 11-58.

- [Habert et al. 1997] B. Habert, A. Nazarenko et A. Salem, *Les linguistiques de corpus*, Paris, Armand Colin.
- [Habert et al. 1998] B. Habert, C. Fabre et F. Issac, *De l'écrit au numérique : constituer, normaliser, exploiter les corpus électroniques*. Paris, InterEditions – Masson.
- [Halliday and Hasan 1976] M.A.K Halliday and R. Hasan, *Cohesion in English*, London, Longman.
- [Harris 1952] Z. Harris, Discourse Analysis, *Language*, Vol. 28, pp. 1-30.
- [Harris 1968] Z. Harris, *Mathematical Structures of Language*, New-York, Wiley & Sons.
- [Harris 1988] Z. Harris, *Language and Information*, New-York, Columbia University Press.
- [Harris et al. 1989] Z. Harris, M. Gottfried, T. Ryckman, P. Mattick, A. Daladier, T. Harris and S. Harris, *The form of information in science: Analysis of an immunology sublanguage*, Preface by Hilary Putnam, Dordrecht/Holland and Boston, Kluwer Academic Publishers.
- [Haskel 1971] P.I. Haskel, Collocations as a measure of stylistic variety, in R.A. Wisbey (ed.), *The computer in literary and linguistic research*, Papers from a Cambridge symposium, Cambridge, Cambridge University Press, pp. 159-168.
- [Hausmann 1979] F.-J. Hausmann, Un dictionnaire des collocations est-il possible ?, *Travaux de littérature et de linguistique de l'Université de Strasbourg*, XVII:1, pp. 187-195.
- [Hausmann 1989] F.-J. Hausmann, Le dictionnaire de collocations, dans F.-J. Hausmann et al. (éds.), *Wörterbücher : ein internationales Handbuch zur Lexicographie*. Dictionnaires, Berlin, De Gruyter, pp. 1010-1019.
- [Hausmann 1997] F.-J. Hausmann, Tout est idiomatique dans les langues, dans M. Martins-Baltar (éd.), *Langages, La Locution entre langues et usages*, Fontenay Saint-Cloud, ENS éditions, pp. 277-290.
- [Hausmann 1999] F.-J. Hausmann, Le dictionnaire de collocations – Critères de son organisation, dans N. Greiner et al. (éds.), *Texte und Kontexte in Sprachen und Kulturen, Festschrift für Jörn Albrecht*, Trier : WVT Wissenschaftlicher Verlag, pp. 121-139.
- [Heid 1992] U. Heid, Décrire les collocations. Deux approches lexicographiques et leur application dans un outil informatisé, *Terminologie et traduction*, Vol. 2(3), pp. 523-544.
- [Heid 1994] U. Heid, On Ways Words Work Together - Topics in Lexical Combinatorics, *Euralex'94 Proceedings*, Amsterdam, pp. 226-257.
- [Hoey 1991] M. Hoey, *Patterns of Lexis in Text*, Oxford, Oxford University Press.
- [Howarth 1998] P. Howarth, The Phraseology of Learners' Academic Writing, in A.P. Cowie (ed.) *Phraseology: Theory, Analysis, and Applications*. Oxford, Clarendon Press, pp. 161-186.
- [Ide et Véronis 1996] N. Ide et J. Véronis, Présentation de la TEI : Text Encoding Initiative, *Cahiers de GUTenberg*, N° 24, pp. 4-10.
- [Illouz et al. 1999] G. Illouz, B. Habert, S. Heiden, P. Lafon, S. Fleury et H. Folch, Maîtriser les déluges de données hétérogènes, dans *Corpus et traitement automatique des langues : pour une réflexion méthodologique*, A. Condamines, C. Fabre et M.-P. Péry-Woodley (éds.), pp. 37-46.

- [Jacobi 1984] D. Jacobi, Du discours scientifique, de sa reformulation et de quelques usages sociaux de la science, *Revue Langue Française*, Vol. 64, pp. 37-51.
- [Jones and Sinclair 1974] S. Jones and J. Sinclair, English lexical collocations, *Cahiers de Lexicologie*, N° 24, pp. 15-61.
- [Kahane and Polguère 2001a] S. Kahane and A. Polguère, Formal foundation of lexical functions, in workshop proceedings: *COLLOCATION: Computational Extraction, Analysis and Exploitation*, 39th Annual Meeting and 10th Conference of the European Chapter of the Association for Computational Linguistics, Toulouse, July 7th 2001, pp. 8-15.
- [Kahane et Polguère 2001b] S. Kahane et A. Polguère, *Un langage formel d'encodage des fonctions lexicales et son application à la modélisation des collocations*, dans B. Daille et G. Williams (coord.), Journée d'étude de l'ATALA « Collocations ». <http://atala.biomath.jussieu.fr/je/010113/Kahane.resume.pdf>
- [Kennedy 1990] G. Kennedy, Collocations: Where grammar and vocabulary teaching meet in S. Anivan (ed), *Language Teaching Methodology for the Nineties*, SAEMEO Regional Language Centre, Singapore, pp. 215-229.
- [Kerbrat-Orecchioni 1980] C. Kerbrat-Orecchioni, *L'énonciation de la subjectivité dans le langage*, Paris, Armand Colin.
- [Kilgarrieff 1992] A. Kilgarrieff, *Polysemy*, dissertation, University of Sussex.
- [Kilgarrieff 2001] A. Kilgarrieff, Comparing Corpora, *International Journal of Corpus Linguistics*, Vol. 6(1), pp. 97-133.
- [Kittredge 1982] R. Kittredge, Variation and Homogeneity of Sublanguages, in R. Kittredge and J. Lehrberger (eds.), *Sublanguages: Studies in Restricted Semantic Domains*, Berlin, New York, Walter de Gruyter, pp. 107-137.
- [Kittredge and Grishman 1986] R. Kittredge and R. Grishman, *Analysing Language in Restricted domains: Sublanguage description and processing*, Hillsdale N.-J., Lawrence Erlbaum Associates.
- [Kjellmer 1984] G. Kjellmer, Some Thoughts on Collocational Distinctiveness, in J. Aarts, W. Meijs (eds.), *Corpus Linguistics: recent advances in the use of computer corpora in English* Language research, Amsterdam, Rodopi, pp. 163-171.
- [Kleiber 1984] G. Kleiber, Dénomination et relations dénominatives, *Langages*, N° 76, *La dénomination*, Paris, Larousse, pp. 77-94.
- [Kleiber 1999] G. Kleiber, Problèmes de sémantique. *La polysémie en question*, Nancy, Presses Universitaires du Septentrion.
- [Kocourek 1991] R. Kocourek, *La langue française de la technique et de la science : Vers une linguistique de la langue savante*, Présentation de Alain Rey, 2^{ème} édition, augmentée, refondue et mise à jour avec une nouvelle bibliographie, Wiesbaden, O. Brandstetter Verlag.
- [Lakoff et Johnson 1985] G. Lakoff et M. Johnson, *Les métaphores dans la vie quotidienne*, Paris, Minuit.

- [Laporte et L'homme 1997] I. Laporte et M.-C. L'homme, Recensement et consignation des combinaisons lexicales en langue de spécialité : Exemple d'application dans le domaine de la pharmacologie cardiovasculaire, *Terminologies nouvelles* Vol. 16, pp. 95-101.
- [Lareau 2002] F. Lareau, *La synthèse de textes comme outil de développement et de vérification de modèles linguistiques formels*, Mémoire de maîtrise, Département de linguistique et traduction, Université de Montréal.
- [Larousse Lexis 1998] Larousse Lexis, *Dictionnaire de la langue française*.
- [Lecomte et Paroubek 1996] J. Lecomte et P. Paroubek, *Le catégoriseur d'Eric Brill. Mise en œuvre de la version entraînée à l'InaLF*, Rapport interne, Nancy, CNRS-INaLF.
- [Lerat 1995] P. Lerat, Les langues spécialisées, *Linguistique nouvelle*, Paris, Presses Universitaires de France.
- [L'homme 2000] M.-C L'homme, Understanding Specialized Lexical Combinations, *Terminology*, Vol. 6(1), pp. 86-110.
- [L'homme 2003] M.-C L'homme, Les combinaisons lexicales spécialisées (CLS): Description lexicographique et intégration aux banques de terminologie, dans *Les collocations : analyse et traitement*, F. Grossmann et A. Tutin (éds.), Travaux et recherches en linguistique appliquée, Amsterdam, de Werelt, pp. 89-103.
- [Ludewig 2001] P. Ludewig, LogoTax - un outil exploratoire pour l'étude de collocations en corpus, dans traitement automatique des langues – TAL, Vol. 42(2), *Traitement automatique des langues et linguistique de corpus*, Paris, Hermès, pp. 623-642.
- [Maingueneau 1996] D. Maingueneau, L'analyse du discours en France aujourd'hui, dans *Le discours : enjeux et perspectives*, numéro spécial Le français dans le monde, Paris, Hachette, pp. 8-16.
- [Malinowski 1923] B. Malinowski, The problem of Meaning in Primitive Languages, in C.J. Ogden and I.A. Richards (eds.), *The meaning of Meaning: A Study of the Influence of Language upon Thought and of the Science of Symbolism*, New York, Harcourt, Brace and World, pp. 296-336.
- [Malrieu et Rastier 2001] D. Malrieu et F. Rastier, Genres et variations morphosyntaxiques, *Traitements automatiques du langage*, Vol. 42(2), 2001, pp. 547-577.
- [Martin 1997a] J. Martin, Du bon usage des corpus dans la recherche sur le discours spécifique, *Revue ASp* Vol. 15-18, pp. 75-83.
- [Martin 1997b] R. Martin, Sur les facteurs du figement lexical, dans M. Martins-Baltar (éd.), *La locution, entre langue et usages*, Fontenay Saint-Cloud, ENS éditions, pp. 291-307.
- [Mayaffre 2002] D. Mayaffre, Les corpus réflexifs : entre architextualité et intertextualité, *Corpus*, Vol. 1(1), pp. 51-70.
- [Mayaffre 2005] D. Mayaffre, Rôle et place des corpus en linguistique : réflexions introductives, P. Vergely (coord.), JETOU 2005 (Journées d'Etudes Toulousaines en Sciences du Langage) « Le rôle et la place des corpus en linguistique », Toulouse, pp. 5-17.

- [McNaught 1992] J. McNaught, Introduction to sublanguage: a tutorial, in H. Thompson (ed.), *Record of the workshop on sublanguage grammar and lexicon acquisition for speech and natural language processing*, Edinburgh, pp. 3-18.
- [Mel'cuk 1988] I. Mel'cuk, Paraphrase et lexique dans la théorie linguistique Sens-Texte : Vingt ans après. *Cahiers de lexicologie*, N° 52(1), pp. 5-50, et N° 53(2), pp. 5-53.
- [Mel'cuk 1997] I. Mel'cuk, *Vers une linguistique Sens-Texte, Leçon inaugurale*. Paris, Collège de France, 78 pages.
<http://www.ling.umontreal.ca/olst/FrEng/melcukColldeFr.pdf>
- [Mel'cuk 1998] I. Mel'cuk, Collocations and Lexical Functions, in A.P. Cowie (ed.), *Phraseology. Theory, Analysis, and Applications*, Oxford, Clarendon Press, pp. 23-53.
- [Mel'cuk 2003] I. Mel'cuk, Collocations : définition, rôle et utilité, dans *Les collocations : analyse et traitement*, F. Grossmann et A. Tutin (éds.), Travaux et recherches en linguistique appliquée, Amsterdam, de Werelt, pp. 23-31.
- [Mel'cuk et al. 1984, 1988, 1992, 1999] I. Mel'cuk, Dictionnaire explicatif et combinatoire du français contemporain, *Recherches lexico-sémantiques*, Vols. I, II, III, IV, Montréal, Presses de l'Université de Montréal.
- [Mellet 2002] S. Mellet, Corpus et recherches linguistiques : introduction, *Corpus*, Vol. 1, pp. 5-13.
- [Meunier 1979] A. Meunier, Points de repères historiques sur la notion de modalité, *DRLAV* 21, pp. 17-24.
- [Meyer and Mackintosh 1996] I. Meyer, K. Mackintosh, The corpus from a terminographer's viewpoint, *International Journal of Corpus Linguistics*, Vol. 2, pp. 257-285.
- [Meynard 2000] I. Meynard, *Internet, Répertoire bilingue de combinaisons lexicales spécialisées français-anglais*, Montréal, Linguattech.
- [Moreno et al. 1996] M.T. Moreno, J.I. Cubero, D. Berner, D. Joel, L.J. Musselman and C. Parker (eds.), *Proceedings of the 6th. International Parasitic Weed Symposium*. Advances in Parasitic Plant Research, Dirección General de Investigación Agraria, Servicio de Publicaciones y Divulgación, Cordoba, Spain.
- [Morris 1938] C.W. Morris, Foundation of the Theory of Signs, *International Encyclopedia of Unified Science*, Vol. 1, University of Chicago Press, Chicago, pp. 77-137.
- [Morris 1964] C.W. Morris, *Signification and significance*, MIT Press, Cambridge, Massachusetts.
- [Morris 1971] C.W. Morris, *Writings on the General Theory of Signs*, The Hague, Mouton, pp. 13-71.
- [Mortchev-Bouveret 1998] M. Mortchev-Bouveret, Approche de la dénomination en langue spécialisée, *Meta* Vol. XLIII(3), pp. 393-410.
- [Mortchev-Bouveret 2004] M. Mortchev-Bouveret, Etude comparative de la phraséologie dans six dictionnaires de langue française, in A. Häcki Buhofer and H. Burger (eds.), *Phraseology*

in Motion, Proceedings zu Europhras Basel 2004, Baltmannsweiler, Schneider Verlag, pp. 289-298.

[Mortureux 1982] M.-F. Mortureux, La vulgarisation, *Langue française*, Vol. 53, Paris, Larousse.

[Mortureux 1985] M.-F. Mortureux, Linguistique et vulgarisation scientifique, *Information sur les sciences sociales*, Vol. 24(4), Londres, Sage Publications, pp. 825-845.

[Mortureux 1995] M.-F. Mortureux, Les vocabulaires scientifiques et techniques, *Les Carnets du Cediscor*, 3, Paris, Presses Sorbonne Nouvelle, pp. 14-25.

[Moskovich 1982] W. Moskovich, What is a sublanguage? The notion of sublanguage in modern Soviet linguistics, in R. Kittredge and J. Lehrberger (eds.), *Sublanguage*. Berlin, de Gruyter, pp. 191-205.

[Müller-Thurau 1985] C.P. Müller-Thurau, *Lexikon der Jugensprache*, Düsseldorf – Wien, Econ.

[Namer 2000] F. Namer, Flemm : Un analyseur Flexionnel du Français à base de règles, *Revue TAL*, Vol. 41(2), pp. 523-548.

[Nazarenko 1998] A. Nazarenko, Compositionnalité, *Revue TAL*, Vol. 39(1), pp. 3-7.

[Palmer 1968] F. R. Palmer, *Selected Papers of J.R. Firth 1952-1959*, London, Longman.

[Partington 1998] A. Partington, *Patterns and Meanings*, Amsterdam, John Benjamins.

[Pearson 1998] J. Pearson, *Terms in Context*, Amsterdam, John Benjamins.

[Pecman 2004a] M. Pecman, Exploitation de la phraséologie scientifique pour les besoins de l'apprentissage des langues, Journée d'étude de l'ATALA « Traitement Automatique des Langues et Apprentissage des Langues », édité par Grenoble, LIDILEM, Université de Grenoble Stendhal, pp. 145-154.

[Pecman 2004b] M. Pecman, L'enjeu de la classification en phraséologie, *EUROPHRAS 2004*, 26-29 août 2004, Université de Bâle, Suisse, Baltmannsweiler, Schneider Hohengehren Verlag, pp. 127-146.

[Pecman 2005] M. Pecman, Les apports possibles de la phraséologie à la didactique des langues étrangères, *Apprentissage des Langues et Systèmes d'Information et de Communication (ALSIC)*, N° 8(1), pp. 109-122.

[Péry-Woodley 1995] M.-P. Péry-Woodley, Quels corpus pour quels traitements automatiques ?, *Traitement Automatique des Langues*, Vol. 36(1-2), pp. 213-232.

[Péry-Woodley 2000] M.-P. Péry-Woodley, *Une pragmatique à fleur de texte : approche en corpus de l'organisation textuelle*. Mémoire d'HDR, Carnets de grammaire, N° 8 (juillet 2000), Université de Toulouse-Le Mirail, ERSS.

[Péry-Woodley and Rebeyrolle 1998] M.-P. Péry-Woodley and J. Rebeyrolle, Domain and genre in sublanguage text: definitional microtexts in three corpora, in A. Rubio, N. Gallardo, R. Castro and A. Tejada (eds.), *First International Conference on Language Resources and Evaluation*, Paris, ELRA, pp. 987-992.

- [Phillips 1985] M. Phillips, *Aspects of Text Structure: An investigation of the lexical Organisation of Text*, Amsterdam, North Holland.
- [Polguère 2000a] A. Polguère, Towards a theoretically-motivated general public dictionary of semantic derivations and collocations for French, *Proceedings of EURALEX'2000*, Stuttgart, pp. 517-527.
- [Polguère 2000b] A. Polguère, Une base de données lexicale du français et ses applications possibles en français, *Revue de Linguistique et de Didactique des Langues*, N° 21, pp. 75-97.
- [Polguère 2003] A. Polguère, Étiquetage sémantique des lexies dans la base de données DiCo, *Revue TAL*, Vol. 44(2), pp. 39-68.
- [Polguère 2005] A. Polguère, *Typologie des entités lexicales d'une base de données explicative et combinatoire*, Journée d'étude de l'ATALA « Interface lexique-grammaire et lexiques syntaxiques et sémantiques », École nationale supérieure des télécommunications (ENST), Paris.
- [Pottier 1974] B. Pottier, *Linguistique générale, théorie et description*, Paris, Klincksieck.
- [Putnam 1975] H. Putnam, The meaning of 'meaning'. *Philosophical papers*, Vol. 2, Cambridge, Cambridge University Press.
- [Putnam 1990] H. Putnam, *Représentation et réalité*, Paris, Le Seuil.
- [Putnam 1994] H. Putnam, *Le réalisme à visage humain*, Paris, Armand Colin.
- [Rastier 1987] F. Rastier, *Sémantique interprétative*, Paris, Presses Universitaires de France, (2^{ème} édition, 1996).
- [Rastier 1989] F. Rastier, *Sens et textualité*, Paris, Hachette.
- [Rastier 1991] F. Rastier, *Sémantique et recherches cognitives*, Paris, Presses Universitaires de France.
- [Rastier 2001] F. Rastier, *Arts et sciences du texte*, Paris, Presses Universitaires de France.
- [Rastier 2005] F. Rastier, *Enjeux épistémologiques de la linguistique de corpus*, Rennes, Presses Universitaires de Rennes, pp. 31-45.
- [Rastier et al. 1994] F. Rastier, M. Cavazza et A. Abeillé, *Sémantique pour l'analyse*, Paris, Masson.
- [Rey 2002] A. Rey, Préface du Grand Robert de la langue française, dans *Le Grand Robert de la langue française, deuxième édition du dictionnaire alphabétique et analogique de la langue française de Paul Robert*, A. Rey (dir.), pp. 19-55.
- [Rey et Chantreau 1989] A. Rey et S. Chantreau, *Dictionnaire des expressions et locutions*, Le Robert (éds.), Collection Les usuels.
- [Saussure 1916] F. de Saussure, *Cours de linguistique générale*, C. Bally et A. Sechehaye avec la collaboration de A. Riedlinger (éds.), Lausanne-Paris, Payot.
- [Schrödinger 1935] E. Schrödinger, Die gegenwärtige Situation in der Quantenmechanik, *Naturwissenschaften* Vol. 23, pp. 807-812; 823-828; 844-849.

- [Schrödinger 1992] E. Schrödinger, *Physique quantique et représentation du monde*, Collection Points-Sciences, Paris, Le Seuil.
- [Sinclair 1987] J. Sinclair, Collocation: A progress report. In R. Steele and T. Threadgold, (eds.), *Language Topics: Essays in Honour of Michael Halliday*, Vol. II, Amsterdam, John Benjamins, pp. 319-331.
- [Sinclair 1991] J. Sinclair, *Corpus, concordance, collocation*, Oxford, Oxford University Press.
- [Sinclair 1996] J. Sinclair, *Preliminary recommendations on Corpus Typology*, Rapport technique, CEE, EAGLES (Expert Advisory Group on Language Engineering Standards).
- [Sinclair et al. 1970, 2002] J. Sinclair, S. Jones and R. Daley, English Lexical Studies, *Report to Office for Scientific and Technical Information (OSTI)* on Project C/LP/08, University of Birmingham.
- [Smadja 1993] F. Smadja, Retrieving Collocations from Text: Xtract, *Computational Linguistics*, Vol. 19(1), pp. 143-177.
- [Smadja and McKeown 1991] F. Smadja and K.R. McKeown, *Using collocations for language generation*, Vol. 7(4), pp. 229-239.
- [Sperber 1996] D. Sperber, *La contagion des idées*, Paris, Odile Jacob.
- [Steinlin et al. 2004] J. Steinlin, S. Kahane, A. Polguère et A. El Ghali, De l'article lexicographique à la modélisation objet du dictionnaire et des liens lexicaux, *Actes de EURALEX'2004*, Lorient, pp. 177-186.
- [Stubbs 1995] M. Stubbs, Collocations and semantic profiles, *Functions of Language*, Vol. 2(1), pp. 23-55.
- [Stubbs 1996] M. Stubbs, *Text and Corpus Analysis*, Oxford, Blackwell.
- [Swales 1990] J. Swales, *Genre analysis*, Cambridge, Cambridge University Press.
- [Telija 1992] V.N. Telija, Lexicographic description of words and collocations: feature-functional model, *Euralex '90*, Barcelona, pp. 315-320.
- [Telija 1994] V.N. Telija, Lexical collocation: denominative and cognitive aspects, *Euralex '94*, Amsterdam, pp. 368-377.
- [Telija et al. 1998] V.N. Telija, N.G. Bragina, E.O. Oparina and I.I. Sandomirskaja, Phraseology as a language of culture: its role in the representation of a cultural mentality, in *Phraseology: Theory, Analysis, and Applications*, A.P. Cowie (ed.), Oxford, pp. 55-75.
- [Tellier 1995, 2002] C. Tellier, *Éléments de syntaxe du français*, Montréal, Presses de l'université de Montréal.
- [Termium] http://www.termium.gc.ca/site/accueil_home_f.html
- [TLFi] Le Trésor de la Langue Française informatisé : <http://atilf.atilf.fr/tlf.htm>
- [Tognini-Bonelli 2001] E. Tognini-Bonelli, *Corpus Linguistics at Work*, Amsterdam, John Benjamins.

- [Tutin 1996] A. Tutin, The formalization of collocations for natural language processing, *Proceedings of COMPLEX'96*, Hungary, Budapest, pp. 243-256.
- [Tutin 1997] A. Tutin, Coder les collocations dans un lexique formel de TALN, *Revue française de linguistique appliquée*, Vol. II, pp. 43-58.
- [Tutin 2004] A. Tutin, Pour une modélisation dynamique des collocations dans les textes, *Actes de EURALEX'2004*, France, Lorient, pp. 207-219.
- [Tutin 2005] A. Tutin, Le dictionnaire de collocations est-il indispensable ? *Revue Française de Linguistique Appliquée. Dictionnaires : nouvelles approches, nouveaux modèles*. Vol. X(2), pp. 31-48.
- [Tutin et Grossmann 2002] A. Tutin et F. Grossmann, Collocations régulières et irrégulières : esquisse de typologie du phénomène collocatif, *Revue Française de Linguistique Appliquée, Lexique : recherches actuelles*, Vol. VII, pp. 7-25.
- [Verlinde et al. 2003] S. Verlinde, T. Selva et J. Binon, Les collocations dans les dictionnaires d'apprentissage : repérage, présentation et accès, dans *Les collocations : analyse et traitement*, F. Grossmann et A. Tutin (éds.), Travaux et recherches en linguistique appliquée, Amsterdam, de Werelt, pp. 105-115.
- [Véronis 2000] J. Véronis, Annotation automatique de corpus : panorama et état de la technique, dans J.-M. Pierrel (éd.), *Ingénierie des langues*, Paris, Hermès, pp. 111-129.
- [Vion 2001] R. Vion, Modalités, modalisations et activités langagières, *Marges linguistiques* – N° 2, MLMS (éd.), pp. 209-231.
<http://marges.linguistiques.free.fr>
- [Voorhees and Harman 1999] E. M. Voorhees and D. Harman, Overview of the Eight Text REtrieval Conference (TREC-8), in *Proceedings of Eight Text REtrieval Conference (TREC-8)*, NIST Special Publication: SP 500-246, pp. 1-24.
- [Wanner 1996] L. Wanner, *Lexical Functions in Lexicography and Natural Language Processing*, Amsterdam, John Benjamins.
- [Williams 1998] G. Williams, Corpus : constitution et contraintes, dans P. Boucher (éd.) *Actes des Ateliers en Morphologie*, Université de Nantes, COLEX (Centre Ouest Lexique), pp. 152-158.
- [Williams 1999a] G. Williams, Looking in before looking out: Internal selection criteria in a corpus of plant biology, Papers in computational Lexicography, *Proceedings of Complex'99*, Hungary, Budapest, pp. 195-204.
- [Williams 1999b] G. Williams, *Les réseaux collocationnels dans la construction et l'exploitation d'un corpus dans le cadre d'une communauté de discours scientifique*, thèse de doctorat soutenue le 15/12/99, à l'Université de Nantes.
- [Williams 2001a] G. Williams, *Les réseaux collocationnels dans la construction et l'exploitation d'un corpus dans le cadre d'une communauté de discours scientifique*, Lille, Presses Universitaires de Septentrion.
- [Williams 2001b] G. Williams, Sur les caractéristiques de la collocation, *TALN*, Tours, pp. 9-16.

[Williams 2002] G. Williams, In search of representativity in specialized corpora: categorisation through collocation, *International journal of Corpus Linguistics*, Vol. 7(1), pp. 43-64.

[Williams 2003] G. Williams, Les collocations et l'école contextualiste britannique, dans *Les collocations : analyse et traitement*, F. Grossmann et A. Tutin (éds.), Travaux et recherches en linguistique appliquée, Amsterdam, de Werelt, pp. 33-44.

[Williams 2005] G. Williams (éd.), *La linguistique de corpus*, Rennes, Presses Universitaires Rennes.

[Wouden 1997] T. van der Wouden, *Negative contexts: Collocation, polarity and multiple negation*, London, Routledge Studies in Germanic Linguistics.

[Zgusta 1980] L. Zgusta, *Theory and method in Lexicography: Western and Non-Western Perspectives*, South Carolina, L. Zgusta (ed.), Columbia, South California, Hornbeam Press.

[Zurek 1991] W.H.Zurek, Decoherence and the transition from quantum to classical, in *Physics today*, Vol. 44(10), pp. 36--44.

Annexes

Présentation

Annexe 1 (Chapitre 3) : données sur la communauté TAL.

Annexe 2 (Chapitre 3) : liste de référencement des articles du corpus.

Annexe 3 (Chapitre 4) : liste des concordances droites et gauches de performance-s.

Annexe 4 (Chapitre 5) : articles lexicographiques de améliorer – amélioration.

Annexe 5 (Chapitre 5) : articles lexicographiques de performance.

Annexe 6 (Chapitre 5) : articles lexicographiques de erreur.

Annexe 7 (Chapitre 5) : articles lexicographiques de problème.

Annexe 8 (Chapitre 5) : articles lexicographiques de évaluer – évaluation.

Annexe 9 (Chapitre 5) : articles lexicographiques de résultat.

Annexe 1 (Chap. 3) : données sur la communauté TAL.

La communauté TAL multiplie les données textuelles et les genres : articles de recherche, articles de conférence, résumés d'articles, résumés de thèses, conférences TALN... Nous présentons ci-dessous les différentes données textuelles que nous aurions pu rassembler.

La revue TAL

L'association ATALA publie depuis 1960 la revue internationale Traitement Automatique des Langues (TAL), anciennement La Traduction Automatique, puis TA Informations.

Les articles

L'article se décompose en un titre, un auteur, un résumé, des mots clef, une introduction, un développement (composé d'autant de partie que nécessaires au propos), une conclusion, des remerciements, une bibliographie.

Les auteurs des articles sont des spécialistes du domaine.

Les articles et descriptions de projet sont soumis à des « consignes de rédaction » et des « normes de présentation » strictes. Ces normes de rédaction et de présentation pour la revue Traitement Automatique des Langues sont celles de son éditeur, Hermès/Lavoisier (<http://tal.e-revues.com/appel.jsp>).

Les notes de lecture

Une note de lecture est le résumé d'un ouvrage récemment publié, en lien avec le domaine.

Les auteurs des articles sont généralement des chercheurs ou leurs doctorants.

Elles sont théoriquement soumises à des « consignes de rédaction » et des « normes de présentation ».

Problème : les notes de lectures sont rédigées par des doctorants, c'est à dire des futurs membres du domaine, mais pas encore reconnus comme tels.

Les comptes rendus (de livres)

Les comptes rendus portent sur des ouvrages récents en lien avec tous les thèmes qui touchent d'une manière ou d'une autre au TAL. L'invitation est ainsi formulée, « Si vous avez lu récemment un livre qui vous a enthousiasmé, ou qui vous a agacé, ou qui vous a tout simplement intéressé, nous vous encourageons vivement à faire profiter de vos commentaires l'ensemble de la communauté en publiant un compte rendu dans la revue. » (<http://www.atala.org/tal/comptes-rendus.html>).

Les comptes rendus sont rédigés par les scientifiques ou leurs doctorants sur recommandations.

Les comptes rendus sont théoriquement soumis à des « consignes de rédaction » et des « normes de présentation ».

Problème : même remarque que ci-dessus, les comptes rendus sont rédigées par des doctorants, c'est à dire des futurs membres du domaine, mais pas encore reconnus comme tels.

Les résumés de thèse

La revue TAL publie également des résumés de thèses récemment soutenues (quelques mois). L'exercice prend la forme d'une synthèse de la recherche sur une page, précédée du nom de l'auteur, du titre du travail, du domaine et de la date de soutenance et de la composition du jury.

L'auteur du résumé est le thésard, jeune diplômé.

Les résumés de thèse sont également soumis à des « consignes de rédaction » et des « normes de présentation ».

Problème : même remarque que ci-dessus, les résumés sont rédigés par des doctorants, c'est à dire des futurs membres du domaine, mais pas encore reconnus comme tels.

Sont publiés également dans la revue : Les sommaires de la revue, les éditoriaux, les introductions au thème de la revue, occasionnellement une rubrique informations, parfois des hommages, des index...

Les Colloques

TALN (Traitements Automatique des Langues Naturelles)

La conférence TALN est née en 1994 dans le cadre du pôle Langage Naturel du GdR-PRC Communication Homme-Machine. Elle a été créée par Philippe Blache, qui s'est chargé de ses trois premières éditions et en a lancé la dynamique. TALN est congrès francophone d'audience internationale sur le traitement automatique des langues. Depuis 1997, le comité d'organisation de TALN tourne chaque année et depuis 1999, cette manifestation se déroule sous l'égide de l'ATALA. Un Comité permanent, créé en 1998, recueille les candidatures pour l'organisation des éditions successives de la conférence, et assure sa pérennité.

La conférence TALN a pour but de faire le point sur les avancées de la recherche dans le domaine TALN tout thème confondu. C'est une sorte d'état des lieux du domaine à un moment m.

Cette conférence comprend généralement des communications scientifiques, des conférences invitées, des séances de démonstration et posters, ainsi que des tutoriels.

RECITAL

Depuis 1999, la conférence RÉCITAL (Rencontre des Étudiants Chercheurs en Informatique pour le Traitement Automatique des Langues) se déroule sous la forme d'une session étudiante de TALN. Ce colloque a donc lieu conjointement au colloque TALN, sur le même thème, mais est réservé aux jeunes docteurs et aux doctorants.

L'objectif de RECITAL est réellement de lancer les jeunes chercheurs. Contrairement au fonctionnement des sélections pour TALN, les doctorants qui soumettent un article à RECITAL connaissent le nom de leurs relecteurs ; ils peuvent donc échanger de façon constructive sur tel ou tel refus.

Problème : même remarque que ci-dessus, les conférences sont rédigées par des doctorants, c'est à dire des futurs membres du domaine, mais pas encore reconnus comme tels.

Les Conférences Associées

Il s'agit de conférences spécialisées sur un thème et qui réunit à ce titre la communauté de thème qui y est associée. Les sessions ont lieu parallèlement à aux colloques TALN et RECITAL.

Elles ont lieu dans à la fin de la conférence générale. Il y en a deux ou trois qui rassemblent les communautés de thème concernées. Les thèmes ne sont pas prédéfinis ou indexés sur les publications de la revue, ils sont établis en fonction des propositions d'articles qui sont effectuées pour la session de l'année en cours.

Les Journées d'Etudes (JE)

L'ATALA organise chaque année depuis 1960 plusieurs journées d'études qui comprennent une série d'exposés, éventuellement accompagnés de « démonstrations ». Elles sont l'occasion de rencontres et de discussions entre différents acteurs du domaine. Chaque journée, placée sous la responsabilité d'un organisateur, est consacrée à un thème particulier touchant aux recherches théoriques ou aux applications en linguistique, en informatique et en TAL (Ex : communication homme-machine, dictionnaires électroniques, indexation automatique, traduction automatique, reconnaissance de la parole, synthèse vocale, etc...). Il y a cinq journées d'études organisées par an.

Les auteurs peuvent être des universitaires, des industriels, des chercheurs ou encore des étudiants.

Les exposés ou démonstrations font parfois l'objet d'un support électronique généralement disponible en ligne pour les JE datant de 1999 et plus : <http://www.atala.org/je/je.html>

Problème : toutes les journées ne font pas l'objet d'une mise en ligne, et d'autre part, ces présentations comportent énormément de tableaux et autres schémas, ce qui rend la quantité de texte restante quasiment inexistante.

Annexe 2 (Chap. 3) : liste de référencement des articles du corpus.

NOM ORIGINEL	NV NOM	AUTEUR(S) : Articles de recherche	TITRE DU TEXTE	TITRE DE LA REVUE	ANNEE	N° DE LA REVUE
03-Antoine-Goulian.txt	AR_0002.txt	Jean-Yves ANTOINE Jérôme GOULIAN	Etude des phénomènes d'extraction en français parlé sur deux corpus de dialogue oral finalisé. Application à la communication orale homme-machine	Linguistique de corpus	2001	Vol. 42, n°2
05-Salmon-alt.txt	AR_0003.txt	Suzanne SALMON-ALT	Entre corpus et théorie : l'annotation (co)référentielle	Linguistique de corpus	2001	Vol. 42, n°2
07-Ilouz-jardino.txt	AR_0004.txt	Gabriel ILLOUZ Michèle JARDINO	Analyse statistique et géométrique de corpus textuels	Linguistique de corpus	2001	Vol. 42, n°2
08-ferret.txt	AR_0005.txt	Olivier FERRET Brigitte GRAU	Utiliser des corpus pour amorcer une analyse thématique	Linguistique de corpus	2001	Vol. 42, n°2
09-Malrieu-Rastier-docd.txt	AR_0006.txt	Denise MALRIEU François RASTIER	Genres et variations morphosyntaxiques	Linguistique de corpus	2001	Vol. 42, n°2
10-Beauvisage.txt	AR_0007.txt	Thomas BEAUVISAGE	Morphosyntaxe et genres textuels. Exploiter des données morphosyntaxiques pour l'étude statistique des genres textuels : application au roman policier	Linguistique de corpus	2001	Vol. 42, n°2
11-Granger et alii.txt	AR_0008.txt	Sylviane GRANGER Anne VANDEVANTER Marie-Josée HAMEL	Analyse de corpus d'apprenants pour l'ELAO basé sur le TAL	Linguistique de corpus	2001	Vol. 42, n°2
12-Ludewig.txt	AR_0009.txt	Petra LUDEWIG	LogoTax : un outillage exploratoire pour l'étude de collocations en corpus	Linguistique de corpus	2001	Vol. 42, n°2
bellot.txt	AR_0011.txt	Patrice BELLOT Marc EL-BEZE	Classification locale non supervisée pour la recherche documentaire	TAL pour la recherche d'information	2000	Vol. 41, n°2
bouillon.txt	AR_0012.txt	Pierrette BOUILLON Cécile FABRE Pascal SEBILLOT Laurence JACQMIN	Apprentissage de ressources lexicales pour l'extension de requêtes	TAL pour la recherche d'information	2000	Vol. 41, n°2
daille.txt	AR_0013.txt	Béatrice DAILLE Jean ROYAUTE Xavier POLANCO	Evaluation d'une plate-forme d'indexation de termes complexes	TAL pour la recherche d'information	2000	Vol. 41, n°2
dal.txt	AR_0014.txt	Georgette DAL Fiammetta NAMER	Génération et analyses automatiques de ressources lexicales construites utilisables en recherches d'informations	TAL pour la recherche d'information	2000	Vol. 41, n°2
dias.txt	AR_0015.txt	Gael HARRY DIAS Sylvie GUILLORE Jean-Claude BASSANO José-Gabriel PEREIRA LOPES	Extraction automatique d'unités lexicales complexes : un enjeu fondamental pour la recherche documentaire	TAL pour la recherche d'information	2000	Vol. 41, n°2
gaussier.txt	AR_0016.txt	Eric GAUSSIER Grégory GREFENSTETTE David HULL Claude ROUX	Recherche d'information en français et traitement automatique des langues	TAL pour la recherche d'information	2000	Vol. 41, n°2

namer.txt	AR_0017.txt	Fiammetta NAMER	FLEMM : un analyseur flexionnel du français à base de règles	TAL pour la recherche d'information	2000	Vol. 41, n°2
rajman.txt	AR_0018.txt	Martin RAJMAN Romaric BESANCON Jean-Cédric CHAPPELLIER	Le modèle DSIR : une approche à base de sémantique distributionnelle pour la recherche documentaire	TAL pour la recherche d'information	2000	Vol. 41, n°2
composTal.txt	AR_0019.txt	Pascal AMSILI Myriam BRAS	DRT et compositionnalité	Compositionnalité	1998	Vol.39,n°1
3Chambreuil.txt	AR_0020.txt	Michel CHAMBREUIL Abdeljabbar BEN GHARBA Pablo GAMALLO OTERO	Variations sur la compositionnalité Montaguienne	Compositionnalité	1998	Vol.39,n°1
4Gayral.txt	AR_0021.txt	Françoise GAYRAL	Créativité du sens en contexte et hypothèse de compositionnalité	Compositionnalité	1998	Vol.39,n°1
2Godart.txt	AR_0022.txt	Béatrice GODART-WENDLING Frédérique ILDEFONSE Jean- Claude PARIENTE Irène ROSIER	Penser le principe de compositionnalité : éléments de réflexion historiques et épistémologiques	Compositionnalité	1998	Vol.39,n°1
7Victorri.txt	AR_0024.txt	Sabine PLOUX Bernard VICTORRI	Constructions d'espaces sémantiques à l'aide de dictionnaires de synonymes	Compositionnalité	1998	Vol.39,n°1
article.txt	AR_0025.txt	Christèle POIRIER Yann MATHET Patrice ENJALBERT	La compositionnalité à l'épreuve des faits, à travers un projet de compréhension automatique de constats d'accidents	Compositionnalité	1998	Vol.39,n°1
GATCOCH5.txt	AR_0026.txt	José COCH	Applications industrielles de la génération : pourquoi et comment	Génération de textes	1998	Vol.39,n°2
G-TAG.txt	AR_0027.txt	Laurence DANLOS	G-TAG : un formalisme lexicalisé pour la génération de textes inspiré de TAG	Génération de textes	1998	Vol.39,n°2
muller.txt	AR_0028.txt	Philippe MULLER Laure SARDA	Représentation de la sémantique des verbes de déplacement transitifs du français	Génération de textes	1998	Vol.39,n°2
tal98.txt	AR_0029.txt	Franck PANAGET	Le générateur de langage naturel de l'agent dialoguant ARTIMIS	Génération de textes	1998	Vol.39,n°2
LexAPol.txt	AR_0030.txt	Alain POLGUERE	Pour un modèle stratifié de la lexicalisation en génération de texte	Génération de textes	1998	Vol.39,n°2
Pugeault99.txt	AR_0031.txt	Florence PUGEAULT Guy LAPALME	Réutilisabilité de ressources de la sémantique lexicale pour la génération de réponses en langage naturel	Génération de textes	1998	Vol.39,n°2
roussarie.txt	AR_0032.txt	Laurent ROUSSARIE	Le problème de la structuration et de la représentation du discours vu sous l'angle de la génération automatique	Génération de textes	1998	Vol.39,n°2

BechetYvon.txt	AR_0034.txt	Frédéric BÉCHET François YVON	Les noms propres en traitement automatique de la parole	Le traitement automatique des noms propres	2000	Vol.41,n°3
Bodenreider	AR_0035.txt	Olivier BODENREIDER Pierre ZWEIGENBAUM	Stratégies d'identification des noms propres à partir de nomenclatures médicales parallèles	Le traitement automatique des noms propres	2000	Vol.41,n°3
DailleMorin.txt	AR_0036.txt	Béatrice DAILLE Emmanuel MORIN	Reconnaissance automatique des noms propres de la langue écrite : Les récentes réalisations	Le traitement automatique des noms propres	2000	Vol.41,n°3
Grass.txt	AR_0037.txt	Thierry GRASS	Typologie et traductibilité des noms propres de l'allemand vers le français	Le traitement automatique des noms propres	2000	Vol.41,n°3
MaurelPitonEggert.txt	AR_0038.txt	Denis MAUREL Odile PITON Elmar EGGERT	Les relations entre noms propres : Lieux et habitants dans le projet Prolex	Le traitement automatique des noms propres	2000	Vol.41,n°3
talsen.txt	AR_0040.txt	Christophe D'ALESSANDRO	33 ans de synthèse de la parole à partir du texte, une promenade sonore (1968-2001)	Synthèse de la parole à partir du texte	2001	Vol.42,n°1
Le projet EULER05.txt	AR_0041.txt	Michel BAGEIN Thierry DUTOIT Nawfal TOUNSI Fabrice MALFRERE Alain RUELLE Dominique WYNSBERGHE	Le projet EULER : vers une synthèse de la parole générique et multilingue	Synthèse de la parole à partir du texte	2001	Vol.42,n°1
article_lia.txt	AR_0042.txt	Frédéric BÉCHET	LIA-PHON : un système complet de phonétisation de textes	Synthèse de la parole à partir du texte	2001	Vol.42,n°1
talfin.txt	AR_0043.txt	Philippe BOULA DE MAREUIL Christophe D'ALESSANDRO Frédéric BEAUGENDRE Anne LACHERET DUJOUR	Une grammaire en tronçons appliquée à la génération de la prosodie	Synthèse de la parole à partir du texte	2001	Vol.42,n°1
AIX.txt	AR_0044.txt	Albert DI CRISTO Philippe DI CRISTO	Syntax : une approche métrique autosegmentale de la prosodie	Synthèse de la parole à partir du texte	2001	Vol.42,n°1
mingus_fips.txt	AR_0045.txt	Piet MERTENS Jean-Philippe GOLDMAN Eric WEHRLI Arnaud GAUDINAT	La synthèse de l'intonation à partir de structures syntaxiques riches	Synthèse de la parole à partir du texte	2001	Vol.42,n°1
morel.txt	AR_0046.txt	Michel MOREL Anne LACHERET DUJOUR	"Kali" : synthèse vocale à partir du texte	Synthèse de la parole à partir du texte	2001	Vol.42,n°1
tal.txt	AR_0048.txt	Philippe BOULA DE MAREUIL Philippe CELERIER T. CESSÉ S. FABRE C. JOBIN Pierre-Yves LE MEUR D. OBADIA B. SOULAGE J. TOEN	ELAN Text-to-Speech : un système multilingue de synthèse de la parole à partir du texte	Synthèse de la parole à partir du texte	2001	Vol.42,n°1
CVOX_TAL.txt	AR_0049.txt	Nicolas TORZEC Thierry MOUDENC Françoise EMERARD	Prétraitements et analyses linguistiques dans le système de parole à partir du texte CVOX : application à la vocalisation automatique d'Emails	Synthèse de la parole à partir du texte	2001	Vol.42,n°1

tal3.txt	AR_0050.txt	Evelyne TZOUKERMANN	La synthèse de la parole du Français : le système des laboratoires Bell	Synthèse de la parole à partir du texte	2001	Vol.42,n°1
KM.txt	AR_0051.txt	Sylvain KAHANE Igor MELCUK	Synthèse des phrases à extraction en français contemporain (du réseau sémantique à l'arbre syntaxique)	Varia	1999	Vol.40,n°2
TAL.txt	AR_0052.txt	Thierry POIBEAU Adeline NAZARENKO	L'extraction d'information, une nouvelle conception de la compréhension de texte ?	Varia	1999	Vol.40,n°2
apb-tal-1999.txt	AR_0053.txt	André POPESCU-BELIS	Evaluation numérique de la résolution de la référence : critiques et proposition	Varia	1999	Vol.40,n°2
tal02_eval.txt	AR_0054.txt	Laurence DEVILLIERS Hélène MAYNARD Patrick PAROUBEK	Méthodologies d'évaluation des systèmes de dialogue parlé : réflexions et expériences autour de la compréhension	Dialogue	2002	Vol.43,n°2
final.txt	AR_0055.txt	Frédéric LANDRAGIN Suzanne SALMONT-ALT Laurent ROMARY	Ancrage référentiel en situation de dialogue	Dialogue	2002	Vol.43,n°2
tal02-vf.txt	AR_0056.txt	Nicolas MAUDET	A la recherche de la structure intentionnelle dans le dialogue	Dialogue	2002	Vol.43,n°2
TAL02.txt	AR_0057.txt	Laurent PREVOST Philippe MULLER Pascale DENIS Laure VIEU	Une approche sémantique et rhétorique du dialogue. Un cas d'étude : l'explication d'un itinéraire	Dialogue	2002	Vol.43,n°2
BrunJacquemSegond.txt	AR_0058.txt	Caroline BRUN Bernard JACQUEMIN Frédérique SEGOND	Exploitation de dictionnaires électroniques pour la désambiguïsation sémantique lexicale	Lexiques sémantiques	2001	Vol.42,n°3
ChiVilFin.txt	AR_0059.txt	Karim CHIBOUT Anne VILNAT Xavier BRIFFAULT	Sémantique du lexique verbal : un modèle en arborescence avec les graphes conceptuels	Lexiques sémantiques	2001	Vol.42,n°3
tal-ls.txt	AR_0060.txt	Vincent CLAVEAU Pascale SEBILLOT Pierrette BOUILLON Cécile FABRE	Acquérir des éléments du Lexique Génératif : quels résultats et à quels coûts?	Lexiques sémantiques	2001	Vol.42,n°3
dailletal01.txt	AR_0061.txt	Béatrice DAILLE	L'identification en corpus d'adjectifs relationnels : une piste linguistique pour l'extraction automatique de terminologie	Lexiques sémantiques	2001	Vol.42,n°3
Article V3.txt	AR_0062.txt	Olivier KRAIF	Exploitation des cognats pour l'alignement : architecture et évaluation	Lexiques sémantiques	2001	Vol.42,n°3
tal99.txt	AR_0063.txt	Emmanuel MORIN	Acquisition de patrons lexico-syntaxiques pour aider au dépouillement terminologique	Multilinguisme	1999	Vol.40,n°1

CanditoOK.txt	AR_0064.txt	Marie-Hélène CANDITO	Un outil multilingue de construction semi-automatique de Grammaire d'Arbres Adjoints. Application au français et à l'italien	Multilinguisme	1999	Vol.40,n°1
TAL-Bourigault-Lame.txt	AR_0065.txt	DidierBOURIGAULT Guiraud LAME	Application à la construction d'une ontologie documentaire du Droit	Structuration de terminologie	2002	Vol.43,n°1
Article-mlv7.txt	AR_0066.txt	Mathieu LAFOURCADE ViolainePRINCE Didier SCHWAB	Vecteurs conceptuels et structuration émergente de terminologies	Structuration de terminologie	2002	Vol.43,n°1
tal-lhomme.txt	AR_0067.txt	Marie-Claude L'HOMME	Fonctions lexicales pour représenter les relations sémantiques entre termes	Structuration de terminologie	2002	Vol.43,n°1
Szulman.txt	AR_0068.txt	Sylvie SZULMAN Brigitte BIEBOW Nathalie AUSSENAC-GILLES	Structuration de terminologies à l'aide d'outils de TAL avec TERMINAE	Structuration de terminologie	2002	Vol.43,n°1
LeRouxOK.txt	AR_0069.txt	Estelle LE ROUX Karine LESPINASSE	Structurer des termes et des informations en langue générale pour décrire des documents audiovisuels	Structuration de terminologie	2002	Vol.43,n°1
art-bes-hermes.txt	AR_0071.txt	Gabriel G. BES	La linguistique entre science et ingénierie	Problèmes épistémologiques	2002	Vol.43,n°3
TAL-V3.txt	AR_0072.txt	Philippe BLACHE	Analyse des dépendances à distance à l'aide de graphes de contraintes. Le cas des disloquées	Problèmes épistémologiques	2002	Vol.43,n°3
CoriLeon.txt	AR_0073.txt	Marcel CORI Jacqueline LEON	La constitution du TAL : Etude historique des dénominations et des concepts	Problèmes épistémologiques	2002	Vol.43,n°3
LavenusTAL.txt	AR_0074.txt	Karine LAVENUS Guy LAPALME	Évaluation des systèmes de question réponse : aspects méthodologiques	Problèmes épistémologiques	2002	Vol.43,n°3
HabertOK.txt	AR_0075.tx	Benoit HABERT Pierre ZWEIGENBAUM	Régler les règles	Problèmes épistémologiques	2002	Vol.43,n°3

NOM ORIGINEL	NV NOM	AUTEUR(S) : Articles de conférence	TITRE DU TEXTE	TITRE DE LA REVUE	ANNEE	LIEU
A1.txt	AC_0001.txt	Claude LAÏ Robert PASERO	Technique de résolution de proformes enchassées	TALN	1999	Cargèse
A11.txt	AC_0002.txt	Patrice LOPEZ Christine FAY VARNIER Azim ROUSSANALY	Sous-langages d'application et LTAG : le système EGAL	TALN	1999	Cargèse
A15.txt	AC_0003.txt	Ronan PICHON Pascale SEBILLOT	Différencier les sens des mots à l'aide du thème et du contexte de leurs occurrences : une expérience	TALN	1999	Cargèse
A23.txt	AC_0004.txt	Françoise GAYRAL Patrick SAINT DIZIER	Peut-on couper à la polysémie verbale?	TALN	1999	Cargèse
A24.txt	AC_0005.txt	Gabriel BES Philippe BLACHE	Propriétés et analyse d'un langage	TALN	1999	Cargèse
A25.txt	AC_0006.txt	Béatrice DAILLE	Identification des adjectifs relationnels en corpus	TALN	1999	Cargèse
A26.txt	AC_0007.txt	Cédric FAIRON Jean SENELLART	Réflexions sur la localisation, l'étiquetage, la reconnaissance et la traduction d'expressions linguistiques complexes	TALN	1999	Cargèse
A29.txt	AC_0008.txt	Olivier KRAIF	Identification des cognats et alignement bi-textuel : une étude empirique	TALN	1999	Cargèse
A32.txt	AC_0009.txt	JeanPhilippe GOLDMAN Christopher LAENZLINGER Eric WEHRLI	La phonétisation de "plus", "tous", et de certains nombres : une analyse phono-syntaxique	TALN	1999	Cargèse
A36.txt	AC_0010.txt	Natalia GRABAR Pierre ZWEIGENBAUM	Acquisition automatique de connaissances morphologiques sur le vocabulaire médical	TALN	1999	Cargèse
A39.txt	AC_0011.txt	Laurent MICLET Jacques CHODOROWSKI	Apprentissage et évaluation de modèles de langage par des techniques de correction d'erreurs	TALN	1999	Cargèse
A40.txt	AC_0012.txt	Georgette DAL Nabil HATHOUT Fiammetta NAMER	Construire un lexique dérivationnel : théorie et réalisations	TALN	1999	Cargèse
A44.txt	AC_0013.txt	Piet MERTENS	Un algorithme pour la génération de l'intonation dans la parole de synthèse	TALN	1999	Cargèse
A5.txt	AC_0014.txt	Béatrice BOUCHOU Denis MAUREL	Une bibliothèque d'opérateurs linguistiques pour la consultation de bases de données en langue naturelle	TALN	1999	Cargèse
A52.txt	AC_0015.txt	Graham WILCOCK	Héritage multiple et templates dans l'implantation de HPSG	TALN	1999	Cargèse

A53.txt	AC_0016.txt	Grabriel ILLOUZ	Méta-étiqueteur adaptif: vers une utilisation pragmatique des ressources linguistiques	TALN	1999	Cargèse
A57.txt	AC_0017.txt	Frédéric MEUNIER	Modélisation des ressources linguistiques d'une application industrielle	TALN	1999	Cargèse
A59.txt	AC_0018.txt	Frédéric BECHET Alexis NASR Thierry SPRIET Renato DE MORI	Modèles de langage à portée variable: application au traitement des homophones	TALN	1999	Cargèse
A62.txt	AC_0019.txt	Christiane PANISSOD	Quantification et anaphore: entité anaphorique complexe (méronymique, processuelle, situationnelle)	TALN	1999	Cargèse
A67.txt	AC_0020.txt	Brahim DJIOUA	DISCC: un outil de construction et d'utilisation d'une base de connaissances sémantico-cognitives des verbes	TALN	1999	Cargèse
A7.txt	AC_0021.txt	Ludovic TANGUY Susan ARMSTRONG Derek WALKER	Isotopies sémantiques et vérification de traduction	TALN	1999	Cargèse
A73.txt	AC_0022.txt	Alexandra KINYON	Hiérarchisation d'analyses basée sur des informations dépendancielles dans le cadre des LTAGs	TALN	1999	Cargèse
A76.txt	AC_0023.txt	Gilles ADDA Joseph MARIANI Patrick PAROUBEK Martin RAJMAN Josette LECOMTE	Métrique et premiers résultats de l'évaluation GRACE des étiqueteurs morpho-syntaxiques pour le français	TALN	1999	Cargèse
A9.txt	AC_0024.txt	Ismaïl BISKRI Sylvain DELISLE	Un modèle hybride pour le textual data mining: un mariage de raison entre le numérique et le linguistique	TALN	1999	Cargèse
art01_p33_42.texte	AC_0025.txt	Anne ABEILLE Lionel CLEMENT Alexandra KINYON François TOUSSENEL	Premières interrogations d'un corpus français arboré	TALN	2001	Tours
art02_p43_52.texte	AC_0026.txt	Tassadit AMGHAR Delphine BATTISTELLI Thierry CHARNOIS	Représenter le temps en langue dans le formalisme graphes conceptuels: une approche basée sur les schèmes sémantico-cognitifs	TALN	2001	Tours
art03_p53_62.txt	AC_0027.txt	Nicolas AUCLERC Yves LEPAGE	Aide à l'analyse pour la construction de banque d'arbres: étude de l'effort	TALN	2001	Tours
art04_p63_72.texte	AC_0028.txt	François BARTHELEMY Pierre BOULIER Philippe DESCHAMP Linda KAOUANE Eric VILLEMONTÉ	Atelier ATOLL pour les grammaires d'arbres adjoints	TALN	2001	Tours
art06_p83_91.texte	AC_0029.txt	Romarc BESANCON Antoine ROZENKNOP Jean-Cédric CHAPPELIER Martin RAJMAN	Intégration probabiliste de sens dans la représentation de textes	TALN	2001	Tours

art07_p93_102.txt	AC_0030.txt	Ismail BISKRI Sylvain DELISLE	Les n-grams de caractères pour l'extraction de connaissances dans les bases de données textuelles multilingues	TALN	2001	Tours
art08_p103_112.texte	AC_0031.txt	Philippe BLACHE	Dépendances à distance dans les grammaires de propriétés: l'exemple des disloquées	TALN	2001	Tours
art09_p113_122.texte	AC_0032.txt	Béatrice BOUCHOU Julien LERAT Denis MAUREL	Une application des classes d'objets: l'interrogation de bases de données	TALN	2001	Tours
art10_p123_132.txt	AC_0033.txt	Estelle CAMPIONE Jean VERONIS	Etiquetage prosodique semi-automatique des corpus oraux	TALN	2001	Tours
art11_p133_142.texte	AC_0034.txt	Jean-Cédric CHAPPELIER Martin RAJMAN	Grammaire à substitution d'arbre de complexité polynomiale: un cadre efficace pour DOP	TALN	2001	Tours
art13_p153_162.texte	AC_0035.txt	Olivier FERRET Brigitte GRAU Martine HURAUULT PLANTET Gabriel ILLOUZ Christian JACQUEMIN	Utilisation des entités nommées et des variantes terminologiques dans un système de question-réponse	TALN	2001	Tours
art14_p163_172.texte	AC_0036.txt	Olivier FERRET Brigitte GRAU Jean-Luc MINEL Sylvie PORHIEL	Repérage des structures thématiques dans des textes	TALN	2001	Tours
art16_p183_192.txt	AC_0037.txt	Nathalie FRIBURGER Denis MAUREL	Elaboration d'une cascade de transducteurs pour l'extraction de motifs: l'exemple des noms de personnes	TALN	2001	Tours
art17_p193_202.texte	AC_0038.txt	Jean-Gabriel GANASCIA	Extraction automatique de motifs syntaxiques	TALN	2001	Tours
art18_p203_212.texte	AC_0039.txt	Jérôme GOULIAN Jean-Yves ANTOINE	Compréhension automatique de la parole combinant syntaxe locale et sémantique globale pour une CHM portant sur des tâches relativement complexes	TALN	2001	Tours
art19_p213_222.texte	AC_0040.txt	Thierry HAMON Adeline NAZARENKO	Exploitation de l'expertise humaine dans un processus de constitution de terminologie	TALN	2001	Tours
art20_p223_232.texte	AC_0041.txt	Nabil HATHOUT	Analogies morpho-synonymiques. Une méthode d'acquisition automatique de liens morphologiques à partir d'un dictionnaire de synonymes	TALN	2001	Tours
art21_p233_242.texte	AC_0042.txt	Mathieu LAFOURCADE Violaine PRINCE	Synonymies et vecteurs conceptuels	TALN	2001	Tours
art22_p243_252.texte	AC_0043.txt	Philippe LANGLAIS Michel SIMARD	Récupération de segments sous-phrastiques dans une mémoire de traduction	TALN	2001	Tours

art23_p253_262.texte	AC_0044.txt	Thomas LEBARBE	Vers une plate-forme multi-agents pour l'exploration et le traitement linguistiques	TALN	2001	Tours
art24_p263_272.texte	AC_0045.txt	Christophe LUC	Une typologie des énumérations basée sur les structures rhétoriques et architecturales du texte	TALN	2001	Tours
art25_p273_282.txt	AC_0046.txt	François MANIEZ	Désambiguïsation syntaxique des groupes nominaux en anglais médical : étude des structures adjectivales à partir d'un corpus bilingue	TALN	2001	Tours
art26_p283_292.txt	AC_0047.txt	Archibald MICHIELS	DEFI: un outil d'aide à la compréhension	TALN	2001	Tours
art27_p293_302.txt	AC_0048.txt	Thierry POIBEAU	Extraction d'information dans les bases de données textuelles en génomique au moyen de transducteurs à nombre fini d'états	TALN	2001	Tours
TALN12.txt	AC_0049.txt	Romarc BESANCON Martin RAJMAN	Filtrages syntaxiques de co-occurrences pour la représentation vectorielle de documents	TALN	2002	Nancy
TALN10.txt	AC_0050.txt	Brigitte BIGI Kamel SMAILI	Identification thématique hiérarchique : application aux forums de discussions	TALN	2002	Nancy
TALN19.txt	AC_0051.txt	Philippe BLACHE Albert DI CRISTO	Variabilité et dépendances des composants linguistiques	TALN	2002	Nancy

Annexe 3 (Chap. 4) : liste des concordances droites et gauches de performance-s.

Concordances performance droit :

NOMFICH	GAUCHE	OCCUR	DROITE
AC_0121	corpus de 500 énoncés ont montré la bonne	performance	à détecter et à corriger les extragrammaticalités ainsi
AR_0013	mise en place pratique et le critère de	performance	appliquée à la fois pour les systèmes d'
AR_0013	indexation libre donne des résultats en termes de	performance	complémentaires de ceux d'une indexation contrôlée.
AR_0035	Nous avons également cherché à lier la	performance	de ces méthodes aux caractéristiques des terminologies,
AR_0035	Afin d'évaluer la	performance	de chaque critère ainsi que celle des combinaisons
AC_0016	coupure/ où /coupure/ donne la	performance	de e j sur le sous-corpus
AR_0017	1996) veulent prouver qu'en anglais la	performance	de l'algorithme employé (troncation vs démarche
AC_0096	Cette combinaison permet effectivement d'augmenter la	performance	de l'analyse syntaxique retournée et plus particulièrement
AC_0063	un modèle de traduction statistique peut améliorer la	performance	de l'engin de traduction sous-jacent
AC_0081	Cette approche leur a permis d'améliorer la	performance	de leur système de 28%.
AC_0094	montré que les meilleurs systèmes testés obtenaient une	performance	de l'ordre de 75% (précision
AR_0046	En définitive, nous pouvons souligner la	performance	de nos méthodes de modification, qui n'
AR_0043	évaluer l'opportunité des frontières intermédiaires et la	performance	de notre analyse syntaxique, nous les avons
AC_0103	Pour améliorer la qualité et la	performance	de notre analyseur, nous avons choisi d'
AC_0063	Dans cette étude, nous évaluons la	performance	de notre moteur de traduction en terme de
AC_0067	personnes pour le français), améliorer la	performance	de notre système de façon importante.
AC_0067	2000, par exemple, indique que la	performance	de notre système était sensiblement inférieure à celle
AC_0063	l'impact du type de corpus sur la	performance	de notre système.
AR_0018	chunking) apporte des améliorations légères de la	performance	de recherche en anglais (Hull D.
AR_0054	la	performance	de reconnaissance d'un système de commande vocale
AC_0010	fiable et peut être utilisé pour améliorer la	performance	des algorithmes de désuffixage.
AR_0018	La	performance	des calculs de similarités peut être augmentée en
AC_0076	à 50 caractères, afin de mesurer la	performance	des deux modules constituant la recherche de la
AC_0046	méthodes de désambiguïsation sont envisageables pour améliorer la	performance	des programmes d'analyse morpho-syntaxique automatique
AR_0018	qui se sont avérées utiles pour augmenter la	performance	des systèmes de recherche documentaire.
AC_0036	Afin d'améliorer la	performance	des systèmes de résumé automatique ou de filtrage
AR_0034	les données traitées entraînera directement une baisse de	performance	des systèmes de TAL, essentiellement pour les
AC_0116	Les	performance	d'étiquetage sur ces catégories s'en ressentira
AR_0048	de synthèse, des outils de mesure de	performance	du code, et des outils si nécessaire
AR_0013	de la chaîne documentaire ILIAD porte sur la	performance	du système à proposer aux indexeurs des descripteurs
AC_0063	Nous quantifions en particulier la baisse de	performance	du système et développons l'idée que l'

AC_0109	globale de cette méthode est meilleure que la	performance	du traitement entraîné sur l'ensemble des textes
AR_0017	hypothèse de départ, à savoir que la	performance	d'un algorithme de racinisation (troncation à
AR_0009	Langer explique ces divergences de	performance	d'un côté par les quantités différentes de
AC_0063	quantifions ensuite en section 3 la baisse de	performance	d'un engin entraîné sur un corpus "
AR_0043	, à partir de l'identification de la	performance	d'un locuteur humain.
AR_0054	" gold standard ") la mesure de	performance	d'un opérateur humain (obtenue avec une
AR_0054	dialogue parlé présente une difficulté supplémentaire : la	performance	d'un opérateur humain sur la tâche est
AR_0036	. S. (1997) évaluent la	performance	d'un programme de reconnaissance d'entités nommées
AR_0054	Oz) et d'exprimer les mesures de	performance	d'un système par rapport à cet étalon
AR_0013	L'évaluation s'intéresse à mesurer la	performance	d'une indexation contrôlée et d'une indexation
AC_0119	un bon système de sélection peut améliorer la	performance	d'une interrogation en langue naturelle.
AC_0117	une décision à 1, la meilleure	performance	en précision mesurée lors de l'évaluation GRACE
AC_0016	Leur	performance	en terme de précision et de décision iv
AC_0034	mauvaises pour les deux modèles : cette mauvaise	performance	est liée à la nature trop bruitée de
AC_0101	Leur	performance	est supérieure à 90% et même à
AR_0018	Toutefois, une légère amélioration de la	performance	est visible lorsqu'on utilise les contextes syntaxiques
AC_0016	ne fournissent pas une homogénéité en termes de	performance	et 3) qu'un même texte peut
AC_0109	2.4 pour augmenter la corrélation entre	performance	et classification, en utilisant une méthode des
AC_0109	traits utilisé pour obtenir une corrélation forte entre	performance	et classification.
AC_0034	2. La	performance	et la précision sur le corpus ATIS corpus
AC_0016	ne fournissent pas une homogénéité en terme de	performance	et un même texte peut produire des variations
AC_0109	(2) la	performance	globale de cette méthode est meilleure que la
AC_0034	et «%correct overall » indique la	performance	globale du modèle, c.-à
AC_0034	La différence en	performance	globale observée est effectivement statistiquement significative à un
AC_0119	que la fusion par le score présente une	performance	intéressante et similaire à la meilleure approche étudiée
AC_0119	Nous pensons que ces différences de	performance	laissent penser qu'un bon système de sélection
AC_0119	par le score ne détériore que peu la	performance	moyenne (perte moyenne de 4.61
AC_0119	une alternative très intéressante au niveau de la	performance	moyenne dont la différence n'est pas significative
AC_0016	Pour ce faire, la	performance	moyenne du MEA sur des textes nouveaux doit
AC_0119	Les résultats démontrent que la	performance	moyenne est très souvent plus élevée dans chacune
AC_0119	Selon nos expériences, la	performance	moyenne présente une baisse moyenne de 12.
AC_0109	variations de performances (par rapport à sa	performance	moyenne) positives pour un, négatives pour
AC_0119	document pertinent) permet d'accroître légèrement la	performance	moyenne.
AR_0071	qu'elle n'est concernée ni par la	performance	ni par l'acquisition du langage, mais

AC_0049	Du fait du faible gain en	performance	observé dans nos expériences, il apparaît que
AC_0063	utilisées, nous pensons que la différence de	performance	observée dans ces deux études s'explique par
AC_0119	Dans nos expériences, la	performance	obtenue est légèrement inférieure à celle obtenue en
AC_0117	Les mesures de	performance	ont été effectuées sur 20 000 mots pour
AC_0117	d'abord été normalisé puis des mesures de	performance	ont été réalisées à l'aide des
AR_0036	l'on souhaite obtenir un indice élevé de	performance	ou une catégorisation fine (McDonald D.
AR_0012	l'expérience E et de la mesure de	performance	P, si sa capacité à exécuter la
AC_0116	La seconde ligne indique le gain de	performance	par rapport à "T_{base}"
AC_0016	plupart des systèmes obtiennent peu de variations de	performance	pour le texte 6, ce qui le
AR_0013	Cette	performance	prend en compte le bruit et le silence
AC_0117	relecture de seulement 10En ce qui concerne les	performance	pure, le taux d'erreur résiduel est
AR_0012	//coupure/définir une mesure de	performance	qui permettra à l'algorithme d'apprentissage de
AR_0013	première, appelée Moyenne orientée utilisateur reflète la	performance	qu'un utilisateur peut attendre du système et
AR_0054	utilisées ne représentent pas fidèlement le domaine de	performance	réelle d'un humain, c'est pourquoi
AR_0013	Le critère de	performance	reflète l'efficacité du système et évalue à
AC_0109	Dans (DeRose, 1988), des	performance	s différentes sont obtenues pour un seul étiqueteur
AC_0116	La baisse de	performance	se produisant pour des valeurs de "S_d
AC_0016	Y a-t-il variation de	performance	selon la situation ?
AR_0046	retombées pratiques en termes d'efficacité et de	performance	sont indiscutables : le fichier de règles s'
AR_0054	à parcourir pour l'amener au niveau de	performance	souhaité,
AC_0117	résiduel est aussi situé dans une plage de	performance	supérieure ou égale aux meilleures performances des systèmes
AC_0043	La	performance	sur le bitexte Verne est moindre, mais
AC_0109	(1) la	performance	sur une partition est meilleure avec le traitement
AC_0119	1. le modèle probabiliste Okapi présente une	performance	très attractive;

Concordances performance gauche :

NOMFICH	GAUCHE	OCCUR	DROITE
AR_0012	-ci, la mesure de leur "	performance	".
AC_0121	corpus de 500 énoncés ont montré la bonne	performance	à détecter et à corriger les extragrammaticalités ainsi
AC_0119	car cette approche a déjà prouvée sa bonne	performance	(Michie et al. 1994) et
AR_0013	Cette	performance	prend en compte le bruit et le silence
AC_0067	qui nous a permis d'arriver à cette	performance	.
AC_0116	La baisse de	performance	se produisant pour des valeurs de "S_d
AC_0063	quantifions ensuite en section 3 la baisse de	performance	d'un engin entraîné sur un corpus "
AC_0063	3.3 Analyse de la baisse de	performance	
AC_0063	Nous quantifions en particulier la baisse de	performance	du système et développons l'idée que l'
AR_0034	les données traitées entraînera directement une baisse de	performance	des systèmes de TAL, essentiellement pour les
AR_0013	Le critère de	performance	reflète l'efficacité du système et évalue à
AR_0013	mise en place pratique et le critère de	performance	appliquée à la fois pour les systèmes d'
AC_0063	utilisées, nous pensons que la différence de	performance	observée dans ces deux études s'explique par
AC_0119	Nous pensons que ces différences de	performance	laissent penser qu'un bon système de sélection
AR_0009	Langer explique ces divergences de	performance	d'un côté par les quantités différentes de
AR_0054	utilisées ne représentent pas fidèlement le domaine de	performance	réelle d'un humain, c'est pourquoi
AR_0036	l'on souhaite obtenir un indice élevé de	performance	ou une catégorisation fine (McDonald D.
AR_0046	retombées pratiques en termes d'efficacité et de	performance	sont indiscutables : le fichier de règles s'
AC_0106	, pour des raisons de faisabilité et de	performance	, nous n'appliquons pas la chaîne de
AC_0116	La seconde ligne indique le gain de	performance	par rapport à "T_{base}
AR_0048	de synthèse, des outils de mesure de	performance	du code, et des outils si nécessaire
AR_0054	" gold standard ") la mesure de	performance	d'un opérateur humain (obtenue avec une
AR_0012	l'expérience E et de la mesure de	performance	P, si sa capacité à exécuter la
AR_0012	//coupure/définir une mesure de	performance	qui permettra à l'algorithme d'apprentissage de
AC_0117	d'abord été normalisé puis des mesures de	performance	ont été réalisées à l'aide des
AC_0117	Les mesures de	performance	ont été effectuées sur 20 000 mots pour
AC_0023	également un biais potentiel dans les mesures de	performance	.
AR_0054	Oz) et d'exprimer les mesures de	performance	d'un système par rapport à cet étalon
AR_0054	à parcourir pour l'amener au niveau de	performance	souhaité,
AC_0117	résiduel est aussi situé dans une plage de	performance	supérieure ou égale aux meilleures performances des systèmes
AR_0014	de deux suffixes peut poser des problèmes de	performance	, et que, bien que les dérivés
AR_0043	Les __ structures de	performance	__ de Grosjean (utilisées dans [BAC90

AC_0101	test avec lequel on calcule le taux de	performance	.
AC_0016	ne fournissent pas une homogénéité en terme de	performance	et un même texte peut produire des variations
AR_0052	Les résultats ont été décevants en termes de	performance	: la stratégie mise en oeuvre est trop
AC_0016	ne fournissent pas une homogénéité en termes de	performance	et 3) qu'un même texte peut
AR_0013	indexation libre donne des résultats en termes de	performance	complémentaires de ceux d'une indexation contrôlée.
AC_0016	Y a-t-il variation de	performance	selon la situation ?
AC_0016	plupart des systèmes obtiennent peu de variations de	performance	pour le texte 6, ce qui le
AC_0016	à mettre en relation avec les variations de	performance	(cf. section 4.2)
AC_0016	le jeu de caractères et les variations de	performance	.
AC_0016	le jeu de caractères et les variations de	performance	.
AC_0109	Dans (DeRose, 1988), des	performance	s différentes sont obtenues pour un seul étiqueteur
AC_0034	La différence en	performance	globale observée est effectivement statistiquement significative à un
AC_0049	Du fait du faible gain en	performance	observé dans nos expériences, il apparaît que
AC_0080	de chacune des méthodes dans le gain en	performance	, ce qui permettrait ensuite l'utilisation d'
AC_0109	2.4 pour augmenter la corrélation entre	performance	et classification, en utilisant une méthode des
AC_0109	traits utilisé pour obtenir une corrélation forte entre	performance	et classification.
AR_0018	La	performance	des calculs de similarités peut être augmentée en
AR_0054	la	performance	de reconnaissance d'un système de commande vocale
AC_0043	La	performance	sur le bitexte Verne est moindre, mais
AC_0109	(1) la	performance	sur une partition est meilleure avec le traitement
AC_0109	(2) la	performance	globale de cette méthode est meilleure que la
AC_0119	Selon nos expériences, la	performance	moyenne présente une baisse moyenne de 12.
AC_0119	Dans nos expériences, la	performance	obtenue est légèrement inférieure à celle obtenue en
AC_0016	Pour ce faire, la	performance	moyenne du MEA sur des textes nouveaux doit
AC_0034	2. La	performance	et la précision sur le corpus ATIS corpus
AR_0054	dialogue parlé présente une difficulté supplémentaire : la	performance	d'un opérateur humain sur la tâche est
AC_0067	personnes pour le français), améliorer la	performance	de notre système de façon importante.
AC_0036	Afin d'améliorer la	performance	des systèmes de résumé automatique ou de filtrage
AC_0081	Cette approche leur a permis d'améliorer la	performance	de leur système de 28%.
AC_0119	un bon système de sélection peut améliorer la	performance	d'une interrogation en langue naturelle.
AC_0063	un modèle de traduction statistique peut améliorer la	performance	de l'engin de traduction sous-jacent
AC_0046	méthodes de désambiguïsation sont envisageables pour améliorer la	performance	des programmes d'analyse morpho-syntaxique automatique
AC_0010	fiable et peut être utilisé pour améliorer la	performance	des algorithmes de désuffixage.
AR_0017	1996) veulent prouver qu'en anglais la	performance	de l'algorithme employé (truncation vs démarche

AC_0096	Cette combinaison permet effectivement d'augmenter la	performance	de l'analyse syntaxique retournée et plus particulièrement
AR_0018	qui se sont avérées utiles pour augmenter la	performance	des systèmes de recherche documentaire.
AR_0018	Toutefois, une légère amélioration de la	performance	est visible lorsqu'on utilise les contextes syntaxiques
AR_0043	, à partir de l'identification de la	performance	d'un locuteur humain.
AR_0018	chunking) apporte des améliorations légères de la	performance	de recherche en anglais (Hull D.
AR_0071	les chomskiens nomment le " modèle de la	performance	7 ".
AR_0075	de la compétence vers des modèles de la	performance	.
AC_0119	une alternative très intéressante au niveau de la	performance	moyenne dont la différence n'est pas significative
AR_0014	raison que nous soit apparue relève de la	performance	: nous faisons l'hypothèse que la faible
AC_0016	coupure/ où /coupure/ donne la	performance	de e j sur le sous-corpus
AR_0043	évaluer l'opportunité des frontières intermédiaires et la	performance	de notre analyse syntaxique, nous les avons
AC_0103	Pour améliorer la qualité et la	performance	de notre analyseur, nous avons choisi d'
AR_0036	. S. (1997) évaluent la	performance	d'un programme de reconnaissance d'entités nommées
AR_0035	Afin d'évaluer la	performance	de chaque critère ainsi que celle des combinaisons
AC_0063	Dans cette étude, nous évaluons la	performance	de notre moteur de traduction en terme de
AC_0034	et «%correct overall » indique la	performance	globale du modèle, c.-à
AC_0119	document pertinent) permet d'accroître légèrement la	performance	moyenne.
AR_0035	Nous avons également cherché à lier la	performance	de ces méthodes aux caractéristiques des terminologies,
AR_0013	L'évaluation s'intéresse à mesurer la	performance	d'une indexation contrôlée et d'une indexation
AC_0076	à 50 caractères, afin de mesurer la	performance	des deux modules constituant la recherche de la
AC_0032	pays qui accueille la compétition, ou la	performance	(le rang) d'un skieur dans
AR_0071	qu'elle n'est concernée ni par la	performance	ni par l'acquisition du langage, mais
AC_0119	par le score ne détériore que peu la	performance	moyenne (perte moyenne de 4.61
AC_0119	Les résultats démontrent que la	performance	moyenne est très souvent plus élevée dans chacune
AC_0067	2000, par exemple, indique que la	performance	de notre système était sensiblement inférieure à celle
AC_0109	globale de cette méthode est meilleure que la	performance	du traitement entraîné sur l'ensemble des textes
AR_0017	hypothèse de départ, à savoir que la	performance	d'un algorithme de racinisation (troncation à
AR_0013	première, appelée Moyenne orientée utilisateur reflète la	performance	qu'un utilisateur peut attendre du système et
AR_0046	En définitive, nous pouvons souligner la	performance	de nos méthodes de modification, qui n'
AC_0063	l'impact du type de corpus sur la	performance	de notre système.
AR_0013	de la chaîne documentaire ILIAD porte sur la	performance	du système à proposer aux indexeurs des descripteurs
AC_0116	Les	performance	d'étiquetage sur ces catégories s'en ressentira
AC_0117	relecture de seulement 10En ce qui concerne les	performance	pure, le taux d'erreur résiduel est
AC_0101	Leur	performance	est supérieure à 90% et même à
AC_0016	Leur	performance	en terme de précision et de décision iv
AC_0034	mauvaises pour les deux modèles : cette mauvaise	performance	est liée à la nature trop bruitée de

AC_0117	une décision à 1, la meilleure	performance	en précision mesurée lors de l'évaluation GRACE
AC_0016	de texte, l'étiqueteur présentant la meilleure	performance	, en moyenne, sur ce type est
AC_0096	si les analyseurs syntaxiques choisis sont de même	performance	.
AC_0067	d'analyse plus généraux afin d'améliorer notre	performance	, tels que désambiguïser des, qui peut
AC_0067	nous avons identifiés nous permettra d'améliorer notre	performance	.
AC_0109	variations de performances (par rapport à sa	performance	moyenne) positives pour un, négatives pour
AR_0009	de la forme canonique et pour améliorer sa	performance	.
AC_0094	montré que les meilleurs systèmes testés obtenaient une	performance	de l'ordre de 75% (précision
AC_0119	1. le modèle probabiliste Okapi présente une	performance	très attractive;
AC_0119	que la fusion par le score présente une	performance	intéressante et similaire à la meilleure approche étudiée

Concordances droite performances :

NOMFICH	GAUCHE	OCCUR	DROITE
AC_0054	Notre méthode de détermination des vocabulaires atteint des	performances	3 fois supérieures à celles obtenues avec des
AC_0054	choix des vocabulaires : nos tests donnent des	performances	3 fois supérieures à celles obtenues avec un
AC_0077	d'accentuation présentées ici permettent d'obtenir des	performances	absolues honorables.
AR_0036	processus est réitéré, jusqu'à obtenir des	performances	acceptables.
AR_0034	Les	performances	actuelles des ordinateurs autorisent l'utilisation, par
AR_0034	De surcroît, les	performances	actuelles des systèmes de reconnaissance limitent bien souvent
AR_0034	% de mots les plus fréquents, les	performances	approchent 90% de correction ; si par
AC_0080	Le tableau 2 présente les meilleures	performances	associées à trois des méthodes de détection de
AC_0080	combinons les méthodes avec ces vocabulaires, les	performances	atteignent o?? g , ce qui
AC_0061	Les	performances	atteintes par Nemesis, sur les anthroponymes et
AR_0053	est de pouvoir choisir la plus sensible aux	performances	attendues, augmentant ainsi les variations des scores
AC_0085	comment elle peut être étendue pour améliorer les	performances	au-delà de ce qui a été
AR_0060	en section 5, et en comparant ses	performances	avec celle présentée en section 4, il
AR_0016	avons pu mettre en évidence une amélioration des	performances	avec une indexation plus fine des documents,
AR_0005	souvent d'un écart faible pour obtenir des	performances	beaucoup moins bonnes.
AC_0109	-dire que les traitements spécialisés ont des	performances	comparables ou meilleures que celles du traitement généraliste
AR_0075	teur puisse choisir son module en fonction des	performances	comparées des modules disponibles assurant une tâche donnée
AC_0050	Enfin, la dernière section donne les	performances	concluantes que nous avons obtenues avec notre approche
AC_0068	le corpus d'entraînement permet d'obtenir des	performances	correctes sur le corpus de test (80
AC_0027	créer des analyseurs stochastiques afin de transporter les	performances	d'analyse linguistique, de l'homme,
AR_0018	l'unité linguistique dans la collection améliore les	performances	dans le cadre de la RD (Salton
AC_0109	traitement est tellement spécialisé qu'il améliore les	performances	d'apprentissage en diminuant celles du test.
AC_0080	le modèle unigramme, ce dernier atteignant des	performances	de ??%i ?? g
AC_0050	observer que le modèle unigramme hiérarchique augmente les	performances	de 2 % de rappel et de précision
AR_0034	1999) atteint des	performances	de 90.7% sur les textes
AR_0034	plus résistants au bruit, bien que leurs	performances	de base ne soient pas sensiblement meilleures :
AR_0005	Nous montrerons que les	performances	de ce second processus sont meilleures concernant la
AC_0108	Nous montrerons que les	performances	de ce second processus sont meilleures concernant la
AR_0055	représentation partiellement hiérarchique obtenue permet d'optimiser les	performances	de certaines opérations spatiales, ce qui contribue
AC_0054	Afin de mieux analyser les	performances	de ces 3 méthodes, il serait donc
AR_0053	une idée expérimentale (§6) des	performances	de cette mesure.

AR_0035	Les performances	de cette méthode dépendent d'une part des
AC_0054	Pour cette raison, nous présentons les performances	de chacune des méthodes en termes de rappel
AC_0109	à dire en évaluant pour chaque texte les performances	de chaque traitement et en gardant la meilleure
AC_0101	Pour mesurer les performances	de classifieurs, nous utilisons la méthode standard
AC_0047	ai déjà fait référence dans l'indication des performances	de DEFI. On trouvera le fichier complet
AC_0054	malgré des études prouvant la non amélioration des performances	de détection en utilisant des séquences de mots
AC_0054	et éventuellement des connaissances syntaxiques pour améliorer les performances	de détection.
AC_0049	par les groupes syntaxiques ne change pas les performances	de façon très significative.
AR_0017	en évidence successivement: (1) les performances	de FLEMM, à grande échelle, en
AR_0040	vitesse d'élocution, afin de démontrer les performances	de KALI pour l'intelligibilité en lecture rapide
AR_0062	d'améliorer, sur ce corpus, les performances	de la détermination des cognats : nous montrons
AC_0008	indiqué une méthode permettant d'améliorer les performances	de la détermination des cognats : nous montrons
AC_0054	significativement les performances	de la TFIDF, qui est la méthode
AC_0057	Les variantes sont plus complexes et améliorent peu les performances	de l'algorithme de base, que nous
AC_0084	Cette opération vise à améliorer les performances	de l'analyseur, mais elle permet également
AR_0041	Pour améliorer les performances	de l'analyseur, un lexique de locutions
AR_0012	avons donc effectué plusieurs tentatives et calculé les performances	de l'apprentissage, et notre choix final
AR_0013	documentation massive sur lesquelles on savait que les performances	de l'automatisation seraient bonnes.
AC_0116	Il est intéressant de constater que les performances	de l'étiquetage sont similaires sur les deux
AR_0017	(iii) qui permet d'améliorer les performances	de l'étiquetage, surtout dans le cas
AR_0017	sur les facultés de FLEMM à améliorer les performances	de l'étiqueteur avec lequel il est associé
AR_0065	Les auteurs évaluent les performances	de leur outil de détection de liens de
AR_0016	1997) obtiennent une amélioration des performances	de leur système de l'ordre de 40
AC_0080	permet d'améliorer de façon très significative les performances	de méthodes de détection lorsqu'elles sont combinées
AC_0061	ce processus apporte une amélioration non négligeable aux performances	de Nemesis (+0,6%
AR_0048	français par exemple, nous avons testé les performances	de notre convertisseur graphème-phonème sur un
AC_0054	3, nous comparons les performances	de notre méthode à la TFIDF et au
AC_0116	Ce corpus nous permet d'évaluer les performances	de notre méthode sur des noms propres ayant
AR_0016	sur les éléments de ce type améliore les performances	de notre système de recherche, il n'
AC_0035	De plus, les performances	de notre système sont en général meilleures pour
AC_0061	Pour améliorer les performances	de notre système, nous avons mis au
AC_0079	Kilgarrieff, Rosenzweig, 2000) atteignent des performances	de plus de 80% pour les noms
AR_0005	de la table 3 montrent clairement que les performances	de SEGAPSITH(1) sont meilleures que
AC_0108	par une première méthode de segmentation : les performances	de SEGAPSITH(1), qui est

AR_0053	R. Mitkov (1998) analyse les	performances	de son système en termes d'amélioration par
AR_0012	NV), en vue d'améliorer les	performances	de systèmes de recherche d'information.
AC_0083	Pour représenter les	performances	de tels systèmes en fonction des différentes valeurs
AR_0017	Les	performances	de TREETAGGER en termes de segmentation se mesurent
AC_0054	Nous avons évalué les	performances	de WSIM sur chacun des deux vocabulaires.
AC_0054	Afin d'étudier les	performances	de WSIM, nous proposons de la comparer
AC_0089	Les	performances	de XIP pour la grammaire générale de l'
AR_0075	Les	performances	dé-croissent dans l'ordre suivant :'
AC_0080	, ce qui correspond à une amélioration des	performances	d'environ j g (les performances les
AR_0034	systèmes de reconnaissance de parole, dont les	performances	dépendent crucialement de la connaissance de l'ensemble
AC_0101	5.1 Comparaison des	performances	des algorithmes
AC_0101	complémentaires influencent légèrement les résultats en améliorant les	performances	des algorithmes.
AC_0106	L'amélioration des	performances	des applications de TAL nécessite d'effectuer des
AC_0079	De plus, les	performances	des critères qui tiennent compte de la position
AR_0054	L'écart entre les	performances	des deux systèmes est plus flagrant pour les
AC_0083), ce qui explique les similitudes de	performances	des deux systèmes.
AR_0034	La mesure généralement utilisée pour comparer les	performances	des différents systèmes se fonde sur le score
AR_0075	de styles distincts, entraînait des variations de	performances	des étiqueteurs (Illouz, 1999).
AR_0017	v) ce programme améliore en outre les	performances	des étiqueteurs qui ont catégorisé les corpus avec
AC_0080	cherché les vocabulaires qui permettaient de maximiser les	performances	des méthodes indépendamment les unes des autres.
AC_0054	Nous pouvons remarquer que les	performances	des méthodes varient significativement d'un thème à
AC_0080	cette approche permet non seulement d'améliorer les	performances	des méthodes, mais exploite également des vocabulaires
AC_0049	est cruciale, par son influence sur les	performances	des modèles à base de co-occurrences
AC_0018	Il est normal que les	performances	des modèles M3 et M4 employés seuls soient
AC_0033	Les	performances	des modules existants sont tout à fait honnêtes
AC_0035	Les	performances	des moteurs de recherche que nous avons utilisé
AR_0007	d'entre elles permet d'augmenter sensiblement les	performances	des outils de classement en éliminant le bruit
AR_0052	On constate avec MUC-5 que les	performances	des principaux systèmes sont devenues comparables et que
AR_0016	la racinisation ne dégrade pas non plus les	performances	des systèmes : un examen détaillé des résultats
AC_0117) utilisé par le NIST pour mesurer les	performances	des systèmes 2.
AR_0034	Ceci explique en partie les faibles	performances	des systèmes à base de règles sur de
AC_0117	qui composent le corpus et les variations des	performances	des systèmes d'annotation morpho-syntaxique associées
AR_0012	: les ressources lexicales permettent d'améliorer les	performances	des systèmes de recherche d'information ; ces
AR_0012	L'augmentation des	performances	des systèmes de recherche d'information, principalement

AC_0036	Afin d'améliorer les	performances	des systèmes de résumé automatique ou de filtrage
AR_0034	les solutions utilisées en pratique pour améliorer les	performances	des systèmes de transcription automatique.
AC_0023	un jeu commun, la mesure effective des	performances	des systèmes d'étiquetage peut commencer.
AC_0023	cadre de GRACE pour la mesure quantitative des	performances	des systèmes d'étiquetage.
AC_0080	parole, un des moyens d'améliorer les	performances	des systèmes est de passer par l'adaptation
AR_0016	sont décevants, dans la mesure où les	performances	des systèmes ne sont peu changées.
AC_0117	Les	performances	des systèmes pour la réalisation de la tâche
AR_0054	développement au cours desquelles les évaluations	performances	des systèmes sont primordiales pour diagnostiquer les
AC_0117	objectives des	performances	capacités
AC_0117	plage de performance supérieure ou égale aux meilleures	performances	des systèmes telle que mesurée dans GRACE (
AC_0117	pour produire les données qui serviront mesurer les	performances	des systèmes.
AC_0117	du protocole qui seront utilisés pour évaluer les	performances	des systèmes.
AR_0060	RI pour étendre des index et augmenter les	performances	des systèmes.
AC_0109	de traits a priori plus corrélés avec les	performances	d'étiquetage comme par exemple : les proportions
AC_0050	classifieurs bayésiens de type unigrammes obtiennent les	performances	d'identification thématique, par rapport à des
AR_0054	meilleures	performances	d'interprétation en contexte sur des difficultés locales
AC_0067	En conclusion, le diagnostic des	performances	d'octobre 2001 de notre système sont meilleures
AR_0034	Nous montrons que les	performances	du décodage acoustique (par augmentation de l'
AC_0054	nombre de variantes est susceptible de dégrader les	performances	du modèle cache ont été dépassées, pour
AC_0034	De plus, les	performances	du modèle DOP restreint par le principe de
AR_0054	Pour évaluer de façon expérimentale les	performances	du module de compréhension indiquent que de la
AR_0054	Les	performances	du module de compréhension.
AR_0017	donc une analyse qualitative et diagnostique automatique	performances	du premier, qui est constitué d'un
AR_0015	des	performances	du processus d'acquisition.
AC_0083	analyse dérivationnelle, de façon à optimiser les	performances	du système d'extraction PLI avec expert ;
AC_0083	1997), améliorent les	performances	du système mixte au cours de sa construction
AC_0115	il suffit de trois itérations pour approcher les	performances	du système PROMÉTHÉE comme un système indépendant.
AR_0052	La figure 2 représente l'évolution des	performances	du système sans pour autant apporter toujours beaucoup
AC_0078	La première expérience évalue les	performances	du système, il faudrait néanmoins savoir dans
AR_0054	dans les textes réels - ralentissent notablement les	performances	du système, les secondes au contraire prennent
AC_0049	Pour pouvoir apprécier les très bonnes	performances	du système.
AC_0063	premières offrent un mode d'évaluation global des	performances	du système.
AR_0016	trop restrictif et détériore de façon significative les	performances	du système.
AR_0012	d'entrées terminologiques améliore de manière	performances	du système.
	significative les	performances	du système.
	nécessaire à la traduction des requêtes ralentit les	performances	du système.
	mots, les résultats obtenus par rapport aux	performances	du test du Chi2/coupure/.

AR_0017	ajout d'un module de lemmatisation augmente les	performances	d'un étiqueteur dans une application en recherche
AC_0016	Une approche possible pour améliorer les	performances	d'un étiqueteur est de l'entraîner et
AC_0116	de corpus sur le Web pouvait améliorer les	performances	d'un étiqueteur stochastique.
AC_0063	Dans ce travail, nous évaluons les	performances	d'un moteur de traduction statistique lorsqu'il
AC_0063	qu'on est en droit d'attendre des	performances	d'un moteur de traduction statistique sur une
AR_0075	Il examine les	performances	d'un parseur probabiliste sur huit domaines du
AR_0017	de flexion et dérivation rend non perfectibles les	performances	d'un raciniseur (Savoy), (
AC_0037	étiquetage peut en effet gêner et diminuer les	performances	d'un système [Morin, 1999]
AC_0085	, qu'il est possible d'accroître les	performances	d'un système de désambiguïsation sémantique au-
AR_0054	3.3 Vers une évaluation prédictive des	performances	d'un système de dialogue
AR_0012	/les ressources lexicales permettent d'améliorer les	performances	d'un système de recherche d'information ;
AR_0013	deux indexations, mais on a évalué les	performances	d'un système de recherche d'information exploitant
AR_0016	évaluer l'influence de la racinisation sur les	performances	d'un système de recherche d'information,
AC_0054	Les	performances	d'un système de reconnaissance automatique de la
AC_0063	Notre but étant avant tout de comparer les	performances	d'un traducteur statistique utilisé dans des conditions
AC_0069	Les	performances	d'une GCFG sont bonnes en auto-
AR_0012	Sans la première étape, une évaluation des	performances	empiriques de la méthode d'apprentissage est possible
AR_0054	permis par exemple de constater que si les	performances	en compréhension sont améliorées par la stratégie,
AC_0080	état de l'art permet d'atteindre des	performances	en détection de thème de ?? e
AC_0080	Enfin, elle permet d'améliorer les	performances	en détection de thème de façon très conséquente
AC_0080	sélection de vocabulaire, nous avons étudié les	performances	en détection de thème en fonction du nombre
AC_0080	lesquelles nous travaillons, nous allons exposer les	performances	en détection de thème obtenues par les méthodes
AC_0080	vocabulaire qui lui est propre et que les	performances	en détection de thème pourraient être améliorées si
AC_0080	Les	performances	en détection de thème sont évaluées par rapport
AC_0080	Dans l'optique d'améliorer les	performances	en détection de thème, nous avons ensuite
AC_0080	proposons permet non seulement d'améliorer légèrement les	performances	en détection de thème.
AC_0080	de près ??ip` g des	performances	en détection de thème.
AC_0080	sélection de vocabulaire fixées, nous évaluons les	performances	en détection de thème.
AC_0080	tailles de vocabulaire, nous avons évalué les	performances	en détection de thème.
AC_0054	Les	performances	en détection de thèmes dépassent également celles du
AC_0080	, elle permet non seulement d'améliorer les	performances	en détection de thèmes des méthodes étudiées,
AC_0109	les résultats, nous arrivons à augmenter les	performances	en diminuant la taille d'entraînement.
AC_0080	information mutuelle) fixées, l'évolution des	performances	en fonction de la taille du vocabulaire.
AC_0085	utilisation des CS apporte une amélioration sensible des	performances	en précision moyenne, ce surcroît d'information
AR_0005	mais elles sont plus équilibrées par rapport aux	performances	en précision, ce que l'on peut

AR_0005	Pour être précis, les	performances	en rappel sont plus faibles mais elles sont
AC_0056	y a peu de possibilités d'accroissement des	performances	en RD par utilisation des synonymes si le
AR_0017	par un étiqueteur/analyseur améliore substantiellement les	performances	en recherche documentaire, soit parce qu'elles
AR_0017	d'information participe indirectement à l'amélioration des	performances	en RI ; il s'agit des codes
AC_0056	de même possible d'obtenir une amélioration des	performances	en utilisant un système de désambiguïsation sémantique (
AC_0029	ce qui semble confirmer que l'amélioration des	performances	est effectivement due à la pertinence de l'
AC_0016	L'effet des textes sur les	performances	est très marqué : les textes 1,
AC_0101	Ces systèmes ont de très bonnes	performances	et commencent à être déployés en environnement industriel
AC_0063	mots inconnus a un impact direct sur les	performances	et en particulier, sur la couverture du
AR_0034	particulière pour de nombreuses applications, dont les	performances	et l'acceptabilité dépendent crucialement de la capacité
AR_0013	de pertinence des documents ramenés améliorent nettement	performances	et rendent l'indexation automatique fonctionnellement
	les		supérieure à
AC_0016	, les systèmes P1 et M2 obtiennent des	performances	faibles pour le texte 1 et sur les
AR_0042	qualité des modèles appris et en conséquence les	performances	globales de la reconnaissance.
AC_0041	L'amélioration des	performances	globales de l'appariement et de la classification
AR_0034	bien souvent un facteur primordial pour juger des	performances	globales de l'application.
AC_0063	insérée à un décodeur afin d'améliorer les	performances	globales de traduction.
AR_0034	la perplexité acoustique, l'impact sur les	performances	globales des systèmes est généralement négative.
AC_0109	d'apprentissage et sur celui de test des	performances	globales du traitement généraliste, du traitement reposant
AC_0060	utilise qu'un contexte local, a des	performances	inférieures à des systèmes tels que CWM,
AC_0080	la Parole (RAP) actuels atteignent des	performances	intéressantes dans des applications ciblées.
AC_0078	robust parsing), ces approches présentent des	performances	intéressantes sur des textes libres.
AC_0050	Ce modèle offre des	performances	légèrement supérieures à celles obtenues avec un unigramme
AC_0080	des performances d'environ j g (les	performances	les plus élevées dans ce cas étaient de
AC_0029	présentés en Figure 2 montrent une amélioration des	performances	lors de l'utilisation de la représentation en
AC_0080	la taille du vocabulaire requise pour atteindre les	performances	maximales (facteur variant entre 2 et 6
AC_0080	remarquer que chacune de ces méthodes obtenait des	performances	maximales avec un vocabulaire qui lui était propre
AC_0080	rechercher les vocabulaires qui permettent d'obtenir les	performances	maximales en combinant les méthodes, ces vocabulaires
AC_0080	Dans le tableau 1 nous présentons seulement les	performances	maximales obtenues par association du meilleur vocabulaire et
AC_0016	entraîné pour une situation donnée peut avoir des	performances	médiocres pour d'autres.
AC_0117	de produire une combinaison de systèmes ayant des	performances	meilleures sur l'une des dimensions (précision
AC_0063	, un regard rapide sur la dégradation des	performances	mesurée sur le corpus SNIPER est tellement criante
AC_0078	dialogue homme-machine finalisé, dont les	performances	montrent la pertinence de ces méthodes pour atteindre
AC_0101	Les deux évaluations ont des	performances	moyennes (bien meilleures qu'un classement aléatoire
AC_0016	des étiqueteurs de maturité équivalente (ayant des	performances	moyennes comparables) entraînés sur des types de
AR_0005	analyse réalisant une segmentation des textes avec des	performances	moyennes et sans identification des thèmes abordés dans

AC_0108	analyse réalisant une segmentation des textes avec des	performances	moyennes et sans identification des thèmes abordés dans
AC_0035	Ces résultats montrent que les	performances	ne diminuent pas quand on traite moins de
AC_0054	Histoire, que le modèle cache, ses	performances	ne semblant pas dépendre de la taille d'
AR_0042	En effet, les	performances	ne sont plus uniquement mesurées de manière quantitatives
AC_0030	Les raisons d'aussi bonnes	performances	nous les retrouvons dans les résultats de la
AC_0063	Deux raisons majeures expliquent les piètres	performances	observées sur le corpus SNIPER: la présence
AC_0023	peuvent par exemple augmenter artificiellement les	performances	observées), des progrès certains ont été
AC_0023	différences de		
AC_0023	une meilleure évaluation des différences significatives	performances	observées.
	entre les		
AC_0080	combinaison, nous avons évalué le gain en	performances	obtenu pour les deux types de vocabulaires étudiés
AC_0049	Les	performances	obtenues avec le modèle de représentation DSIR hybride
AC_0080	de sélection de vocabulaire et nous étudierons les	performances	obtenues avec les nouveaux vocabulaires.
AC_0063	Nous montrons en section 5 les	performances	obtenues en implémentant notre approche et discutons en
AC_0083	Une évaluation des	performances	obtenues par ce système et une comparaison aux
AC_0080	Nous avons évalué les	performances	obtenues par ces méthodes lorsqu'elles sont appliquées
AC_0050	En particulier, on peut observer les	performances	obtenues par deux sous-groupes relatifs à
AC_0109	test que cette méthode permet de dépasser les	performances	obtenues par le partitionnement existant.
AC_0023	1]) à l'étude comparative des	performances	obtenues par plusieurs systèmes.
AR_0052	En dépit des	performances	obtenues, cette analyse locale pose question :
AC_0085	système de désambiguïsation sémantique au-delà des	performances	optimales obtenues de manière statique.
AC_0029	montrent que cette représentation n'améliore pas les	performances	par rapport à la représentation directe en mots
AC_0060	connaissances lexicales générales n'améliore souvent pas	performances	par rapport à l'exploitation de la seule
	les		
AR_0075	s'est centré en tout cas sur les	performances	par rapport à un résultat de référence plutôt
AC_0054	Les	performances	par thème de la TFIDF sont représentatives de
AR_0036	ont permis la construction de syst èmes aux	performances	particulièrement intéressantes.
AC_0030	n'est pas dicté par des raisons de	performances	particulières car tel n'est pas notre objectif
AC_0101	néanmoins les classifieurs lexicaux standards et	performances	peu satisfaisantes.
	obtiennent des		
AC_0050	Ces différences importantes de	performances	peuvent s'expliquer en observant la taille de
AC_0080	vocabulaires n'obtenant peut être pas les meilleures	performances	pour chacune des méthodes utilisées seules.
AR_0034	mais aussi la seule tâche à obtenir des	performances	proches de l'excellence (moins de 10
AC_0096	réaliser une combinaison sur des systèmes ayant des	performances	proches.
AC_0063	Les	performances	que nous avons relevées avec notre décodeur sur
AC_0079	lemmatisation n'a permis une augmentation significative	performances	que pour les adjectifs.
	des		

AC_0047	DEFI devrait offrir ici de meilleures	performances	que sur du texte brut, et les
AC_0116	S_d" est élevée, meilleures sont les	performances	rapportées aux seuls noms propres dont les entrées
AR_0034	Il est difficile d'apprécier avec précision les	performances	réelles de ces systèmes d'identification de la
AR_0016	approche se trouve donc grandement diminué par les	performances	réelles des systèmes sur lesquels elle repose.
AR_0034	des entités nommées est une mesure importante des	performances	réelles du système, dans la mesure où
AC_0101	Les	performances	relatives de chaque algorithme sont similaires à celles
AR_0044	de nous fournir une première indication sur les	performances	relatives de notre système.
AR_0013	Leurs	performances	respectives sont donc complémentaires et montrent l'intérêt
AC_0054	par la méthode WSIM, même si leurs	performances	restent cependant très proches.
AC_0080	que pour les trois méthodes étudiées, les	performances	se sont améliorées.
AC_0016	4.2 Observation des variations de	performances	selon situations.
AC_0054	méthode de détection de thèmes, dont les	performances	seront étudiées, et comparées à d'autres
AC_0118	Les mesures de	performances	sont alors établies en comparant le résultat fourni
AR_0049	Leurs	performances	sont alors suffisantes pour permettre le développement des
AC_0056	de Harman (1988)) que les	performances	sont améliorées lorsque l'utilisateur choisit un certain
AC_0101	Dans tous les cas, les	performances	sont assez impressionnantes, avec une moyenne supérieure
AR_0052	Même si les	performances	sont bonnes (jusqu'à 57 % de
AC_0060	Ses	performances	sont également légèrement supérieures à celles obtenues par
AC_0060	sans restriction quant au domaine abordé mais leurs	performances	sont faibles lorsque la structure thématique des textes
AC_0063	est important de noter, que si les	performances	sont meilleures, tous les problèmes ne sont
AC_0080	mots retenus est inférieur ou supérieur, les	performances	sont plus faibles.
AC_0049	que, pour le corpus OFIL, les	performances	sont plutôt légèrement dégradées, mais les coefficients
AC_0109	ensemble de tests, et conclut que les	performances	sont toujours meilleures si la classe sur laquelle
AC_0080	remarquer que l'ensemble des méthodes atteint des	performances	supérieures à eYfhg .
AC_0080	L'ensemble de ces méthodes atteint des	performances	supérieures à eYfhg .
AC_0116	résultats montrent une légère amélioration par rapport aux	performances	sur "T_{ACS}",
AR_0053	1, afin de permettre la comparaison des	performances	sur des textes différents ; (3)
AR_0034	l'ordre de 5 points par rapport aux	performances	sur des textes non-bruités.
AC_0034	miminale-maximale et d'en évaluer les	performances	sur deux corpus différents.
AC_0034	-- évaluer les	performances	sur le corpus de test ; la mesure
AC_0054	Bien que la méthode WSIM obtienne les meilleures	performances	sur le corpus général, on peut remarquer
AC_0050	Nous avons également montré que même si les	performances	sur le noeud exact ne sont pas élevées
AC_0076	Figure 2 montre que QALC a de meilleures	performances	sur les catégories définies par une entité nommée
AR_0005	de ces paramètres afin de garantir de bonnes	performances	sur un corpus donné.
AR_0053	en général une ou plusieurs mesures de ses	performances	sur un ensemble de données, suivies d'

AC_0056		Les performances	sur une requête qui contiendrait shape pourraient être
AC_0108	les caractéristiques sur un exemple et montrons les	performances	sur une tâche "classique" d'évaluation
AC_0080	Malgré des	performances	très élevées de ce dernier, son utilisation
AC_0117	des algorithmes de résolution différent mais ayant des	performances	voisines (pour le moment il s'agit
AC_0109	faibles différences des résultats et le niveau des	performances) rendent difficile l'utilisation des tests statistiques

Concordances gauche performances :

NOMFICH	GAUCHE	OCCUR	DROITE
AR_0036	ont permis la construction de syst èmes aux	performances	particulièrement intéressantes.
AC_0061	ce processus apporte une amélioration non négligeable aux	performances	de Nemesis (+0,6%
AC_0116	résultats montrent une légère amélioration par rapport aux	performances	sur "T_{ACS}",
AR_0005	mais elles sont plus équilibrées par rapport aux	performances	en précision, ce que l'on peut
AR_0012	mots, les résultats obtenus par rapport aux	performances	du test du Chi2/coupure/.
AR_0034	l'ordre de 5 points par rapport aux	performances	sur des textes non-bruités.
AR_0053	est de pouvoir choisir la plus sensible aux	performances	attendues, augmentant ainsi les variations des scores
AC_0030	Les raisons d'auusi bonnes	performances	nous les retrouvons dans les résultats de la
AR_0005	de ces paramètres afin de garantir de bonnes	performances	sur un corpus donné.
AC_0023	En plus de ses bonnes	performances	, cette méthode présente également l'avantage d'
AC_0101	Ces systèmes ont de très bonnes	performances	et commencent à être déployés en environnement industriel
AC_0078	Pour pouvoir apprécier les très bonnes	performances	du système, il faudrait néanmoins savoir dans
AC_0016	8 et 3 sont cause de baisse de	performances	.
AC_0023	peuvent par exemple augmenter artificiellement les différences de	performances	observées), des progrès certains ont été
AC_0050	Ces différences importantes de	performances	peuvent s'expliquer en observant la taille de
AC_0118	Les mesures de	performances	sont alors établies en comparant le résultat fourni
AC_0030	n'est pas dicté par des raisons de	performances	particulières car tel n'est pas notre objectif
AC_0002	système XTAG, essentiellement pour des raisons de	performances	: plutôt que de reléguer la sélection des
AC_0083), ce qui explique les similitudes de	performances	des deux systèmes.
AR_0048	prosodique est fondé sur les " structures de	performances	" de Grosjean, reprises par [BAC
AR_0005	un point de référence minimal en termes de	performances	, nous avons implémenté une méthode de segmentation
AR_0075	de styles distincts, entraînait des variations de	performances	des étiqueteurs (Illouz, 1999).
AC_0109	qu'un texte peut entraîner des variations de	performances	(par rapport à sa performance moyenne)
AC_0016	4.2 Observation des variations de	performances	selon situations.
AR_0053	une idée expérimentale (§6) des	performances	de cette mesure.
AC_0049	une forte diminution (statistiquement significative) des	performances	.
AC_0060	utilise qu'un contexte local, a des	performances	inférieures à des systèmes tels que CWM,
AC_0056	y a peu de possibilités d'accroissement des	performances	en RD par utilisation des synonymes si le
AC_0054	deux approches n'ont montré aucune amélioration des	performances	.
AC_0041	L'amélioration des	performances	globales de l'appariement et de la classification
AC_0106	L'amélioration des	performances	des applications de TAL nécessite d'effectuer des

AR_0017	d'information participe indirectement à l'amélioration des	performances	en RI ; il s'agit des codes
AC_0029	ce qui semble confirmer que l'amélioration des	performances	est effectivement due à la pertinence de l'
AC_0054	malgré des études prouvant la non amélioration des	performances	de détection en utilisant des séquences de mots
AC_0080	, ce qui correspond à une amélioration des	performances	d'environ j g (les performances les
AC_0049	dont la suppression permet donc une amélioration des	performances	.
AR_0016	avons pu mettre en évidence une amélioration des	performances	avec une indexation plus fine des documents,
AC_0029	présentés en Figure 2 montrent une amélioration des	performances	lors de l'utilisation de la représentation en
AC_0029	tâche de recherche documentaire montrent une amélioration des	performances	, mais une évaluation plus poussée devrait être
AC_0056	de même possible d'obtenir une amélioration des	performances	en utilisant un système de désambiguïsation sémantique (
AR_0016	1997) obtiennent une amélioration des	performances	de leur système de l'ordre de 40
AC_0080	le modèle unigramme, ce dernier atteignant des	performances	de ??%i ?? g
AC_0079	Kilgarriff, Rosenzweig, 2000) atteignent des	performances	de plus de 80% pour les noms
AC_0080	la Parole (RAP) actuels atteignent des	performances	intéressantes dans des applications ciblées.
AC_0080	état de l'art permet d'atteindre des	performances	en détection de thème de ?? e
AR_0034	1999) atteint des	performances	de 90.7% sur les textes
AC_0080	L'ensemble de ces méthodes atteint des	performances	supérieures à eYfhg .
AC_0080	remarquer que l'ensemble des méthodes atteint des	performances	supérieures à eYfhg .
AC_0054	Notre méthode de détermination des vocabulaires atteint des	performances	3 fois supérieures à celles obtenues avec des
AC_0063	qu'on est en droit d'attendre des	performances	d'un moteur de traduction statistique sur une
AR_0012	L'augmentation des	performances	des systèmes de recherche d'information, principalement
AR_0054	donc une analyse qualitative et diagnostique automatique des	performances	du module de compréhension.
AC_0108	analyse réalisant une segmentation des textes avec des	performances	moyennes et sans identification des thèmes abordés dans
AR_0005	analyse réalisant une segmentation des textes avec des	performances	moyennes et sans identification des thèmes abordés dans
AC_0016	entraîné pour une situation donnée peut avoir des	performances	médiocres pour d'autres.
AC_0016	des étiqueteurs de maturité équivalente (ayant des	performances	moyennes comparables) entraînés sur des types de
AC_0117	des algorithmes de résolution différent mais ayant des	performances	voisines (pour le moment il s'agit
AC_0117	de produire une combinaison de systèmes ayant des	performances	meilleures sur l'une des dimensions (précision
AC_0096	réaliser une combinaison sur des systèmes ayant des	performances	proches.
AC_0079	mot à désambiguïser entraîne également une baisse des	performances	.
AR_0062	on n'observe pas une dégradation catastrophique des	performances	: les points d'ancrages issus du préalignement
AC_0101	5.1 Comparaison des	performances	des algorithmes
AC_0083	3.2 Évaluation et comparaison des	performances	
AR_0053	1, afin de permettre la comparaison des	performances	sur des textes différents ; (3)

AC_0023	1)) à l'étude comparative des	performances	obtenues par plusieurs systèmes.
AR_0013	MEDLARS est aussi un facteur de dégradation des	performances	.
AC_0063	, un regard rapide sur la dégradation des	performances	mesurée sur le corpus SNIPER est tellement criante
AC_0049	information et mène donc à une dégradation des	performances	.
AC_0085	système de désambiguïsation sémantique au-delà des	performances	optimales obtenues de manière statique.
AR_0052	En dépit des	performances	obtenues, cette analyse locale pose question :
AR_0054	En conclusion, le diagnostic des	performances	d'interprétation en contexte sur des difficultés locales
AC_0049	de sélection trop fortes entraîne une diminution des	performances	.
AC_0054	choix des vocabulaires : nos tests donnent des	performances	3 fois supérieures à celles obtenues avec un
AC_0023	un jeu commun, la mesure effective des	performances	des systèmes d'étiquetage peut commencer.
AC_0049	Pour ce qui est des	performances	, les résultats de la recherche obtenus pour
AC_0083	Une évaluation des	performances	obtenues par ce système et une comparaison aux
AR_0012	Sans la première étape, une évaluation des	performances	empiriques de la méthode d'apprentissage est possible
AC_0080	information mutuelle) fixées, l'évolution des	performances	en fonction de la taille du vocabulaire.
AC_0083	La figure 2 représente l'évolution des	performances	du système mixte au cours de sa construction
AR_0075	teur puisse choisir son module en fonction des	performances	comparées des modules disponibles assurant une tâche donnée
AC_0080	de près ??ip` g des	performances	en détection de thème.
AR_0054	premières offrent un mode d'évaluation global des	performances	du système, les secondes au contraire prennent
AR_0034	des entités nommées est une mesure importante des	performances	réelles du système, dans la mesure où
AC_0047	ai déjà fait référence dans l'indication des	performances	de DEFI. On trouvera le fichier complet
AR_0034	bien souvent un facteur primordial pour juger des	performances	globales de l'application.
AC_0080	Malgré des	performances	très élevées de ce dernier, son utilisation
AC_0060	Pour la mesure des	performances	, nous avons utilisé la mesure d'erreur
AC_0109	faibles différences des résultats et le niveau des	performances) rendent difficile l'utilisation des tests statistiques
AR_0054	développement au cours desquelles les évaluations	performances	des systèmes sont primordiales pour diagnostiquer les
AR_0054	objectives des	performances	capacités
AR_0054	évaluation et ainsi fournir des mesures objectives des	performances	.
AC_0080	remarquer que chacune de ces méthodes obtenait des	performances	maximales avec un vocabulaire qui lui était propre
AR_0036	processus est réitéré, jusqu'à obtenir des	performances	acceptables.
AR_0034	mais aussi la seule tâche à obtenir des	performances	proches de l'excellence (moins de 10
AC_0068	le corpus d'entraînement permet d'obtenir des	performances	correctes sur le corpus de test (80
AC_0077	d'accentuation présentées ici permettent d'obtenir des	performances	absolues honorables.
AR_0005	souvent d'un écart faible pour obtenir des	performances	beaucoup moins bonnes.
AC_0101	néanmoins les classifieurs lexicaux standards et	performances	peu satisfaisantes.
AC_0016	obtiennent des	performances	.
AC_0016	, les systèmes P1 et M2 obtiennent des	performances	faibles pour le texte 1 et sur les

AC_0050	Ce modèle offre des	performances	légèrement supérieures à celles obtenues avec un unigramme
AC_0101	Les deux évaluations ont des	performances	moyennes (bien meilleures qu'un classement aléatoire
AC_0109	-dire que les traitements spécialisés ont des	performances	comparables ou meilleures que celles du traitement généraliste
AR_0054	3.3 Vers une évaluation prédictive des	performances	d'un système de dialogue
AC_0078	robust parsing), ces approches présentent des	performances	intéressantes sur des textes libres.
AC_0023	3. La mesure quantitative des	performances	: Précision et Décision
AC_0023	cadre de GRACE pour la mesure quantitative des	performances	des systèmes d'étiquetage.
AR_0015	Les résultats obtenus montrent une amélioration sensible des	performances	.
AC_0085	utilisation des CS apporte une amélioration sensible des	performances	en précision moyenne, ce surcroît d'information
AR_0035	aux indices internes apporte une amélioration significative des	performances	.
AC_0079	lemmatisation n'a permis une augmentation significative des	performances	que pour les adjectifs.
AC_0079	grammaticaux se traduit par une baisse significative des	performances	.
AC_0049	le système ne subit aucune dégradation significative des	performances	.
AC_0016	alors répétée en boucle jusqu'à stabilisation des	performances	.
AC_0109	d'apprentissage et sur celui de test des	performances	globales du traitement généraliste, du traitement reposant
AC_0117	qui composent le corpus et les variations des	performances	des systèmes d'annotation morpho-syntaxique associées
AC_0080	combinaison, nous avons évalué le gain en	performances	obtenu pour les deux types de vocabulaires étudiés
AC_0116	récoltées nous donne quelques explications sur ces faibles	performances	: d'une part, même après la
AR_0034	Ceci explique en partie les faibles	performances	des systèmes à base de règles sur de
AC_0054	Les	performances	en détection de thèmes dépassent également celles du
AC_0054	Les	performances	d'un système de reconnaissance automatique de la
AC_0117	Les	performances	des systèmes pour la réalisation de la tâche
AC_0089	Les	performances	de XIP pour la grammaire générale de l'
AR_0034	Les	performances	actuelles des ordinateurs autorisent l'utilisation, par
AC_0049	Les	performances	obtenues avec le modèle de représentation DSIR hybride
AC_0056	Les	performances	sur une requête qui contiendrait shape pourraient être
AC_0033	Les	performances	des modules existants sont tout à fait honnêtes
AC_0054	Les	performances	par thème de la TFIDF sont représentatives de
AC_0061	Les	performances	atteintes par Nemesis, sur les anthroponymes et
AC_0054	Les	performances	de la TFIDF, qui est la méthode
AC_0080	Les	performances	en détection de thème sont évaluées par rapport
AR_0035	Les	performances	de cette méthode dépendent d'une part des
AC_0101	Les	performances	relatives de chaque algorithme sont similaires à celles

AC_0069	Les	performances	d'une GCFG sont bonnes en auto-
AC_0063	Les	performances	que nous avons relevées avec notre décodeur sur
AC_0035	Les	performances	des moteurs de recherche que nous avons utilisé
AR_0075	Les	performances	dé-croissent dans l'ordre suivant :'
AR_0017	Les	performances	de TREETAGGER en termes de segmentation se mesurent
AR_0054	Les	performances	du module de compréhension indiquent que de la
AC_0080	des performances d'environ j g (les	performances	les plus élevées dans ce cas étaient de
AR_0017	en évidence successivement: (1) les	performances	de FLEMM, à grande échelle, en
AC_0101	Dans tous les cas, les	performances	sont assez impressionnantes, avec une moyenne supérieure
AR_0062	d'améliorer, sur ce corpus, les	performances	de la détermination des cognats : nous montrons
AR_0042	En effet, les	performances	ne sont plus uniquement mesurées de manière quantitatives
AC_0080	que pour les trois méthodes étudiées, les	performances	se sont améliorées.
AR_0034	% de mots les plus fréquents, les	performances	approchent 90% de correction ; si par
AC_0049	que, pour le corpus OFIL, les	performances	sont plutôt légèrement dégradées, mais les coefficients
AC_0054	De plus, les	performances	du modèle cache ont été dépassées, pour
AC_0035	De plus, les	performances	de notre système sont en général meilleures pour
AC_0079	De plus, les	performances	des critères qui tiennent compte de la position
AR_0005	Pour être précis, les	performances	en rappel sont plus faibles mais elles sont
AC_0080	mots retenus est inférieur ou supérieur, les	performances	sont plus faibles.
AR_0034	De surcroît, les	performances	actuelles des systèmes de reconnaissance limitent bien souvent
AC_0080	combinons les méthodes avec ces vocabulaires, les	performances	atteignent o?? g , ce qui
AC_0108	par une première méthode de segmentation : les	performances	de SEGAPSITH(1), qui est
AC_0063	Nous montrons en section 5 les	performances	obtenues en implémentant notre approche et discutons en
AC_0085	, qu'il est possible d'accroître les	performances	d'un système de désambiguïsation sémantique au-
AC_0101	complémentaires influencent légèrement les résultats en améliorant les	performances	des algorithmes.
AR_0018	l'unité linguistique dans la collection améliore les	performances	dans le cadre de la RD (Salton
AC_0109	traitement est tellement spécialisé qu'il améliore les	performances	d'apprentissage en diminuant celles du test.
AR_0016	sur les éléments de ce type améliore les	performances	de notre système de recherche, il n'
AR_0015	1997), améliorent les	performances	du processus d'acquisition.
AR_0017	sur les facultés de FLEMM à améliorer les	performances	de l'étiqueteur avec lequel il est associé
AC_0084	Cette opération vise à améliorer les	performances	de l'analyseur, mais elle permet également
AC_0080	Afin d'améliorer les	performances	, nous envisageons d'exploiter les avantages de
AC_0063	insérée à un décodeur afin d'améliorer les	performances	globales de traduction.
AC_0036	Afin d'améliorer les	performances	des systèmes de résumé automatique ou de filtrage
AC_0080	parole, un des moyens d'améliorer les	performances	des systèmes est de passer par l'adaptation

AC_0080	Dans l'optique d'améliorer les	performances	en détection de thème, nous avons ensuite
AR_0017	(iii) qui permet d'améliorer les	performances	de l'étiquetage, surtout dans le cas
AC_0080	Enfin, elle permet d'améliorer les	performances	en détection de thème de façon très conséquente
AR_0012	/les ressources lexicales permettent d'améliorer les	performances	d'un système de recherche d'information ;
AR_0012	: les ressources lexicales permettent d'améliorer les	performances	des systèmes de recherche d'information ; ces
AC_0080	cette approche permet non seulement d'améliorer les	performances	des méthodes, mais exploite également des vocabulaires
AC_0080	, elle permet non seulement d'améliorer les	performances	en détection de thèmes des méthodes étudiées,
AR_0012	NV), en vue d'améliorer les	performances	de systèmes de recherche d'information.
AC_0109	, de façon itérative, en améliorer les	performances	.
AC_0010	requête, elles permettent d'en améliorer les	performances	.
AR_0017	associé (car il peut en améliorer les	performances	: détection et correction d'erreurs d'étiquetage
AC_0061	Pour améliorer les	performances	de notre système, nous avons mis au
AR_0041	Pour améliorer les	performances	de l'analyseur, un lexique de locutions
AC_0085	comment elle peut être étendue pour améliorer les	performances	au-delà de ce qui a été
AC_0016	Une approche possible pour améliorer les	performances	d'un étiqueteur est de l'entraîner et
AR_0034	les solutions utilisées en pratique pour améliorer les	performances	des systèmes de transcription automatique.
AC_0054	et éventuellement des connaissances syntaxiques pour améliorer les	performances	de détection.
AC_0109	aux types de données traitées pour améliorer les	performances	.
AC_0116	de corpus sur le Web pouvait améliorer les	performances	d'un étiqueteur stochastique.
AR_0053	R. Mitkov (1998) analyse les	performances	de son système en termes d'amélioration par
AC_0054	Afin de mieux analyser les	performances	de ces 3 méthodes, il serait donc
AC_0083	il suffit de trois itérations pour approcher les	performances	du système d'extraction PLI avec expert ;
AC_0080	la taille du vocabulaire requise pour atteindre les	performances	maximales (facteur variant entre 2 et 6
AC_0050	observer que le modèle unigramme hiérarchique augmente les	performances	de 2 % de rappel et de précision
AR_0017	ajout d'un module de lemmatisation augmente les	performances	d'un étiqueteur dans une application en recherche
AC_0109	les résultats, nous arrivons à augmenter les	performances	en diminuant la taille d'entraînement.
AR_0060	RI pour étendre des index et augmenter les	performances	des systèmes.
AR_0012	nominal constitue une piste nouvelle pour augmenter les	performances	, et nous nous appuyons sur le Lexique
AC_0016	pour valider une typologie en corrélation avec les	performances	.
AC_0109	de traits a priori plus corrélés avec les	performances	d'étiquetage comme par exemple : les proportions
AR_0012	avons donc effectué plusieurs tentatives et calculé les	performances	de l'apprentissage, et notre choix final
AC_0063	Notre but étant avant tout de comparer les	performances	d'un traducteur statistique utilisé dans des conditions
AR_0034	La mesure généralement utilisée pour comparer les	performances	des différents systèmes se fonde sur le score
AC_0054	3, nous comparons les	performances	de notre méthode à la TFIDF et au

AR_0042	qualité des modèles appris et en conséquence les	performances	globales de la reconnaissance.
AR_0034	nombre de variantes est susceptible de dégrader les	performances	du décodage acoustique (par augmentation de l'
AR_0040	vitesse d'élocution, afin de démontrer les	performances	de KALI pour l'intelligibilité en lecture rapide
AC_0109	test que cette méthode permet de dépasser les	performances	obtenues par le partitionnement existant.
AC_0037	étiquetage peut en effet gêner et diminuer les	performances	d'un système [Morin, 1999]
AC_0050	Enfin, la dernière section donne les	performances	concluantes que nous avons obtenues avec notre approche
AR_0034	particulière pour de nombreuses applications, dont les	performances	et l'acceptabilité dépendent crucialement de la capacité
AC_0078	dialogue homme-machine finalisé, dont les	performances	montrent la pertinence de ces méthodes pour atteindre
AR_0034	systèmes de reconnaissance de parole, dont les	performances	dépendent crucialement de la connaissance de l'ensemble
AC_0054	méthode de détection de thèmes, dont les	performances	seront étudiées, et comparées à d'autres
AC_0054	Nous comparons également les	performances	(en termes de rappel) de notre
AC_0109	Cela fournit une piste pour améliorer encore les	performances	.
AR_0054	L'écart entre les	performances	des deux systèmes est plus flagrant pour les
AC_0023	une meilleure évaluation des différences significatives entre les	performances	observées.
AC_0080	sélection de vocabulaire, nous avons étudié les	performances	en détection de thème en fonction du nombre
AC_0054	Afin d'étudier les	performances	de WSIM, nous proposons de la comparer
AC_0080	de sélection de vocabulaire et nous étudierons les	performances	obtenues avec les nouveaux vocabulaires.
AR_0013	deux indexations, mais on a évalué les	performances	d'un système de recherche d'information exploitant
AC_0054	Nous avons évalué les	performances	de WSIM sur chacun des deux vocabulaires.
AC_0080	tailles de vocabulaire, nous avons évalué les	performances	en détection de thème.
AC_0080	Nous avons évalué les	performances	obtenues par ces méthodes lorsqu'elles sont appliquées
AC_0115	La première expérience évalue les	performances	du système PROMÉTHÉE comme un système indépendant.
AR_0065	Les auteurs évaluent les	performances	de leur outil de détection de liens de
AC_0034	-- évaluer les	performances	sur le corpus de test ; la mesure
AC_0116	Ce corpus nous permet d'évaluer les	performances	de notre méthode sur des noms propres ayant
AC_0034	miminale-maximale et d'en évaluer les	performances	sur deux corpus différents.
AC_0117	du protocole qui seront utilisés pour évaluer les	performances	des systèmes.
AC_0080	sélection de vocabulaire fixées, nous évaluons les	performances	en détection de thème.
AC_0063	Dans ce travail, nous évaluons les	performances	d'un moteur de traduction statistique lorsqu'il
AR_0075	Il examine les	performances	d'un parseur probabiliste sur huit domaines du
AC_0034	Pour évaluer de façon expérimentale les	performances	du modèle DOP restreint par le principe de
AC_0080	lesquelles nous travaillons, nous allons exposer les	performances	en détection de thème obtenues par les méthodes
AC_0080	proposons permet non seulement d'améliorer légèrement les	performances	en détection de thème.
AC_0080	proposé ici a permis d'améliorer légèrement les	performances	(modèle unigramme : ??%i

AC_0101	Ces informations complémentaires améliorent très légèrement les	performances	.
AC_0080	cherché les vocabulaires qui permettaient de maximiser les	performances	des méthodes indépendamment les unes des autres.
AC_0101	Pour mesurer les	performances	de classifieurs, nous utilisons la méthode standard
AC_0117) utilisé par le NIST pour mesurer les	performances	des systèmes 2.
AC_0117	pour produire les données qui serviront mesurer les	performances	des systèmes.
AC_0108	les caractéristiques sur un exemple et montrons les	performances	sur une tâche "classique" d'évaluation
AR_0013	de pertinence des documents ramenés améliorent nettement les	performances	et rendent l'indexation automatique fonctionnellement supérieure à
AC_0100	avéré que ceux-ci dégradent notablement les	performances	.
AR_0052	dans les textes réels - ralentissent notablement les	performances	du système sans pour autant apporter toujours beaucoup
AC_0050	En particulier, on peut observer les	performances	obtenues par deux sous-groupes relatifs à
AC_0080	rechercher les vocabulaires qui permettent d'obtenir les	performances	maximales en combinant les méthodes, ces vocabulaires
AR_0017	analyse dérivationnelle, de façon à optimiser les	performances	du premier, qui est constitué d'un
AR_0055	représentation partiellement hiérarchique obtenue permet d'optimiser les	performances	de certaines opérations spatiales, ce qui contribue
AR_0016	sont décevants, dans la mesure où les	performances	des systèmes ne sont peu changées.
AR_0017	v) ce programme améliore en outre les	performances	des étiqueteurs qui ont catégorisé les corpus avec
AR_0016	approche se trouve donc grandement diminué par les	performances	réelles des systèmes sur lesquels elle repose.
AC_0049	occurrences dont la disparition n'altère pas les	performances	, mais également de filtrer des co-
AC_0029	montrent que cette représentation n'améliore pas les	performances	par rapport à la représentation directe en mots
AC_0049	par les groupes syntaxiques ne change pas les	performances	de façon très significative.
AC_0060	connaissances lexicales générales n'améliore souvent pas les	performances	par rapport à l'exploitation de la seule
AR_0017	de flexion et dérivation rend non perfectibles les	performances	d'un raciniseur (Savoy), (
AC_0057	variantes sont plus complexes et améliorent peu les	performances	de l'algorithme de base, que nous
AR_0016	la racinisation ne dégrade pas non plus les	performances	des systèmes : un examen détaillé des résultats
AR_0034	Il est difficile d'apprécier avec précision les	performances	réelles de ces systèmes d'identification de la
AC_0054	Pour cette raison, nous présentons les	performances	de chacune des méthodes en termes de rappel
AC_0056	de Harman (1988)) que les	performances	sont améliorées lorsque l'utilisateur choisit un certain
AR_0052	On constate avec MUC-5 que les	performances	des principaux systèmes sont devenues comparables et que
AR_0005	de la table 3 montrent clairement que les	performances	de SEGAPSITH(1) sont meilleures que
AC_0109	ensemble de tests, et conclut que les	performances	sont toujours meilleures si la classe sur laquelle
AC_0116	Il est intéressant de constater que les	performances	de l'étiquetage sont similaires sur les deux
AC_0080	vocabulaire qui lui est propre et que les	performances	en détection de thème pourraient être améliorées si
AC_0035	Ces résultats montrent que les	performances	ne diminuent pas quand on traite moins de
AR_0005	Nous montrerons que les	performances	de ce second processus sont meilleures concernant la

AC_0108	Nous montrerons que les	performances	de ce second processus sont meilleures concernant la
AC_0067	Nous montrons que les	performances	d'octobre 2001 de notre système sont meilleures
AC_0018	Il est normal que les	performances	des modèles M3 et M4 employés seuls soient
AC_0054	Nous pouvons remarquer que les	performances	des méthodes varient significativement d'un thème à
AR_0013	documentation massive sur lesquelles on savait que les	performances	de l'automatisation seraient bonnes.
AR_0016	nécessaire à la traduction des requêtes ralentit les	performances	du système.
AC_0083	Pour représenter les	performances	de tels systèmes en fonction des différentes valeurs
AR_0007	d'entre elles permet d'augmenter sensiblement les	performances	des outils de classement en éliminant le bruit
AC_0080	Dans le tableau 1 nous présentons seulement les	performances	maximales obtenues par association du meilleur vocabulaire et
AR_0052	Même si les	performances	sont bonnes (jusqu'à 57 % de
AC_0050	Nous avons également montré que même si les	performances	sur le noeud exact ne sont pas élevées
AC_0063	est important de noter, que si les	performances	sont meilleures, tous les problèmes ne sont
AR_0054	permis par exemple de constater que si les	performances	en compréhension sont améliorées par la stratégie,
AC_0049	trop restrictif et détériore de façon significative les	performances	du système.
AC_0063	d'entrées terminologiques améliore de manière significative les	performances	du système.
AC_0080	permet d'améliorer de façon très significative les	performances	de méthodes de détection lorsqu'elles sont combinées
AC_0008	indiqué une méthode permettant d'améliorer significativement les	performances	de la détermination des cognats : nous montrons
AR_0035	INVren a) ne modifie pas significativement les	performances	.
AC_0116	S_d" est élevée, meilleures sont les	performances	rapportées aux seuls noms propres dont les entrées
AR_0017	par un étiqueteur/analyseur améliore substantiellement les	performances	en recherche documentaire, soit parce qu'elles
AR_0075	s'est centré en tout cas sur les	performances	par rapport à un résultat de référence plutôt
AC_0063	mots inconnus a un impact direct sur les	performances	et en particulier, sur la couverture du
AR_0034	la perplexité acoustique, l'impact sur les	performances	globales des systèmes est généralement négative.
AC_0056	quels enrichissements auront un impact important sur les	performances	.
AR_0044	de nous fournir une première indication sur les	performances	relatives de notre système.
AC_0080	-ci ayant une grande influence sur les	performances	.
AC_0049	est cruciale, par son influence sur les	performances	des modèles à base de co-occurrences
AR_0016	évaluer l'influence de la racinisation sur les	performances	d'un système de recherche d'information,
AC_0016	mesure de l'influence des textes sur les	performances	.
AC_0016	L'effet des textes sur les	performances	est très marqué : les textes 1,
AR_0048	français par exemple, nous avons testé les	performances	de notre convertisseur graphème-phonème sur un
AC_0109	à dire en évaluant pour chaque texte les	performances	de chaque traitement et en gardant la meilleure
AC_0027	créer des analyseurs stochastiques afin de transporter les	performances	d'analyse linguistique, de l'homme,
AR_0049	Leurs	performances	sont alors suffisantes pour permettre le développement des

AR_0013	Leurs	performances	respectives sont donc complémentaires et montrent l'intérêt
AC_0106	sémantiques dans les applications de TAL améliore leurs	performances	.
AR_0074	d'imagination, les uns pour augmenter leurs	performances	, les autres pour suivre l'état de
AC_0060	sans restriction quant au domaine abordé mais leurs	performances	sont faibles lorsque la structure thématique des textes
AR_0034	plus résistants au bruit, bien que leurs	performances	de base ne soient pas sensiblement meilleures :
AC_0054	par la méthode WSIM, même si leurs	performances	restent cependant très proches.
AR_0034	ne produiront pas les mêmes effets sur leurs	performances	.
AC_0117	plage de performance supérieure ou égale aux meilleures	performances	des systèmes telle que mesurée dans GRACE (
AC_0076	Figure 2 montre que QALC a de meilleures	performances	sur les catégories définies par une entité nommée
AC_0047	DEFI devrait offrir ici de meilleures	performances	que sur du texte brut, et les
AC_0054	Bien que la méthode WSIM obtienne les meilleures	performances	sur le corpus général, on peut remarquer
AC_0050	classifieurs bayésiens de type unigrammes obtiennent les meilleures	performances	d'identification thématique, par rapport à des
AC_0080	vocabulaires n'obtenant peut être pas les meilleures	performances	pour chacune des méthodes utilisées seules.
AC_0080	Le tableau 2 présente les meilleures	performances	associées à trois des méthodes de détection de
AC_0117	système composite résultant a toujours eu de meilleurs	performances	(en précision) que n'importe lequel
AC_0063	Deux raisons majeures expliquent les piètres	performances	observées sur le corpus SNIPER: la présence
AC_0060	Ses	performances	sont également légèrement supérieures à celles obtenues par
AC_0054	Histoire, que le modèle cache, ses	performances	ne semblant pas dépendre de la taille d'
AR_0060	en section 5, et en comparant ses	performances	avec celle présentée en section 4, il
AR_0053	en général une ou plusieurs mesures de ses	performances	sur un ensemble de données, suivies d'

Annexe 4 (Chap. 5) : articles lexicographiques de améliorer – amélioration.

Article lexicographique extrait du Larousse Lexis :

AMÉLIORABLE, AMÉLIORATEUR, AMÉLIORATIF, AMÉLIORATION, AMÉLIORER → MEILLEUR.

MEILLEUR, E [mejœr] adj. (lat. *melior*, compar. de *bonus*, bon; 1080). [Av. ou apr. le nom.] 1. Comparatif de supériorité de l'adj. *bon* : *L'hiver a été dur mais le printemps sera meilleur* (Hébert). *Le repas est meilleur qu'hier* (= sa qualité est supérieure). *Ce vin est bien meilleur. Nous avons l'espoir d'un monde meilleur* (= plus juste). *Je vous souhaite une meilleure santé.* — 2. Superlatif : a) accompagné de l'article défini ou indéfini, d'un possessif ou d'un démonstratif et s'accordant avec le mot qualifié : *En taule la pitance littéraire est pour moi la meilleure* (Sarrazin). *Les plaisanteries les plus courtes sont les meilleures.* Cette information est puisée aux meilleures sources. *Je vous présente mes meilleurs vœux, mes souhaits les meilleurs.* *C'est une femme du meilleur monde* (= de la haute société). *J'en passe et des meilleurs* (= je ne parle pas d'autres aventures extraordinaires qui ont eu lieu); b) accompagné de l'article défini et restant invariable : *C'était une façon de proclamer que tout était pour le mieux dans le meilleur des mondes* (Giono). *Dans le meilleur des cas je n'aurais pas pu vous la rapporter* (Faye). • LOC. ADV. et PRÉP. *À meilleur marché*, d'une façon moins coûteuse : *On trouve des asperges à meilleur marché chez l'épicier de la rue de Vaugirard.* || *De meilleure heure*, syn. litt. de PLUS TÔT : *Je me suis levé de meilleure heure pour prendre le train du Havre.* || *De meilleure grâce*, plus spontanément, sans se faire prier. ♦ adv. (1873). *Il fait meilleur, on est mieux.* *Suivant la saison, il fait meilleur dedans ou dehors; il vaut mieux.* *Il fait meilleur espérer que se souvenir; le temps est plus beau.* *En été, il fait meilleur qu'en hiver.* ♦ n. m. (1273). 1. Accompagné de l'article défini invariable, ce qui est excellent dans quelqu'un ou dans quelque chose : *Il lui a consacré le meilleur de sa vie.* *C'est peut-être le meilleur de notre destin d'écrire que ce fil qui court dans la trame d'une vie et la relie à nous* (Mauriac). *Ils sont unis pour le meilleur et pour le pire.* — 2. Prendre le meilleur sur quelqu'un, remporter un avantage sur lui : *Un coureur cycliste qui, au sprint, prend le meilleur sur ses adversaires.* ♦ n. (1080). *Le meilleur, la meilleure, personne supérieure aux autres : Que le meilleur gagne. Même les meilleurs annoncent.* ♦ **améliorer** [ameljore] v. tr. (v. 1160). *Améliorer quelqu'un, quelque chose, le rendre meilleur : Je propose d'y faire de l'élevage en grand et d'améliorer la race bovine* (Giono). *Depuis toujours ils essayaient de m'améliorer, ils n'arrivaient pas à grand-chose* (Céline). *L'augmentation des salaires a amélioré le sort des mineurs* (contr. AGGRAVER). *Améliorer les circuits de distribution* (contr. PERFECTIONNER). ♦ **s'améliorer** v. pr. Devenir meilleur : *Le réseau international s'est amélioré ces dernières semaines* (contr. SE DÉTERIORER). *Son caractère ne s'améliore guère. Ma santé ne s'est pas améliorée* (syn. ALLER MIEUX). ♦ **améliorant, e** adj. (v. 1850). Agr. *Plante améliorante*, celle dont la culture sur un terrain appauvri lui rend sa fertilité et en augmente le rendement par apport d'azote au sol : *Les légumineuses fourragères (luzerne, fèverole, sainfoin) sont des plantes améliorantes.* ♦ **améliorable** adj. (v. 1850). ♦ **amélioration** n. f. (v. 1420, rare jusqu'à 1620). 1. *L'amélioration des rapports entre les deux États a amené une détente générale. Les améliorations apportées à l'immeuble en ont augmenté la valeur* (syn. EMBELLISSEMENT; contr. DÉGRADATION). *Le pauvre homme souffre si fort qu'il a résolu d'attendre dans le néant une amélioration de son état* (Aymé). — 2. Techn. *Amélioration des bois*, ensemble des techniques de traitement des bois, qui aboutissent à des matériaux, les *bois améliorés*, offrant des caractéristiques physiques et mécaniques nouvelles. ♦ **améliorateur, trice** ou **amélioratif, ive** adj. (1866). *Propre à améliorer : Une influence améliorative.*

Article lexicographique extrait du Grand Robert :

AMÉLIORER [ameljɔʁe] v. fr. — 1154, *ameillorer*, de *ā*, et *meillor* « meilleur » (du lat. *meillor*, même sens), reflat en *améliorer* (1507) d'après le lat. *melior*.

♦ 1 Rendre meilleur, plus satisfaisant, changer, transformer en mieux. *Améliorer un produit. Améliorer un mécanisme, un moteur.* → **Perfectionner**. — *Améliorer le fonctionnement, l'organisation de qqch. Améliorer l'ordinaire* (la nourriture, à la caserne). *Améliorer sa santé* (surtout au passif ou au pron.). → **Affermir, développer, fortifier, rétablir. Améliorer un travail, un texte, une traduction.** → **Corriger, retoucher, réviser, revoir. Améliorer les mœurs.** → **Amender, épurer, réformer, régénérer, rénover. Améliorer le sort de qqn.** → **Adoucir. Améliorer les relations entre des personnes, des peuples.** → **Détendre.** *Le service de ce restaurant pourrait facilement être amélioré. Améliorer les routes, les télécommunications.*

(...) instruire le peuple, c'est l'améliorer; éclairer le peuple c'est le moraliser; letter le peuple, c'est le civiliser.

HUGO, *Littérature et Philosophie mêlées*, p. 54.

Ils (les hommes) ont toujours cherché à améliorer leur état.

FRANCE, *le Crime de S. Bonnard*, p. 427.

Tirer le profit le meilleur de ce qui est, singénier à l'améliorer plutôt qu'à chercher à le changer.

GIDE, *Journal*, 4 juil. 1933.

Améliorer une loi par des amendements. → **Amender.**

(Morale). *Vouloir améliorer la société, l'homme, le peuple. Améliorer sa conduite.*

vx. *Améliorer un métal, l'épurer.*

♦ 2 Agric. Rendre (un sol) plus fertile. *Améliorer une terre.* → **Amender, bonifier, fertiliser.**

♦ 3 Apporter des améliorations à (un lieu, une maison); augmenter la valeur d'un bien, d'un immeuble. → **Embellir, réparer, restaurer.**

♦ **S'AMÉLIORER** v. pron. Devenir meilleur. → **Mieux** (aller de mieux en mieux). *Sa santé s'améliore. Les relations se sont améliorées. Le temps s'améliore.* → **Éclaircir (s), embellir.** *Ce vin s'améliore avec les années.* → **Abonir (s), bonifier (se), faire (se). *La situation commence à s'améliorer; ne s'est pas améliorée.***

(...) ces hommes, à mesure que leur sort s'améliorait, sentaient plus amèrement ce qu'il leur restait d'inégalité.

FUSTEL DE COULANGES, *la Cité antique*, p. 313.

(Personnes). *Il ne s'est pas amélioré avec l'âge. S'améliorer intellectuellement, moralement.*

♦ **AMÉLIORÉ, ÉE** p. p. adj. (de *améliorer*). Devenu meilleur. *Des hommes améliorés par l'éducation. Terres améliorées.* → **Amendé.**

CONTR. *Abâtardir, adulterer, aggraver, avarier, corrompre, décliner, dégénérer, dégrader, dépraver, détériorer, empirer, endommager, gâter, pervertir, vicier.* ♦ **DÉR.**

Améliorable, améliorant, améliorateur. — V. **Amélioration, améliorissement.**

AMÉLIORATION [ameljɔʁasjɔ̃] n. f. — 1421; de l'anc. franç. *ameillorer* « améliorer » (d'après lat. *melior*), et *-ation*.

♦ 1 Action de rendre meilleur, de changer en mieux (*l'amélioration de qqch., une amélioration*); fait de devenir meilleur, plus satisfaisant (*une, des améliorations*). → **Changement** (en mieux), **mieux** (n. m.), **perfectionnement, progrès.** *L'amélioration de qqch., dans qqch. Amélioration des qualités d'un produit. L'homme est une machine susceptible d'améliorations.* → **Perfectibilité**, cit. 2. *Légère, nette amélioration du temps.* → **Éclaircie, embellie.** *Aucune amélioration du temps en perspective. Amélioration de l'environnement : mise en œuvre de moyens pour augmenter sa qualité. Amélioration de la santé.* → **Affermissement, rétablissement, Amélioration d'un travail.** → **Correction, retouche, révision. Amélioration des lois, des mœurs.** → **Amendement, réforme, rénovation. Amélioration du sort des affligés.** → **Adoucissement. Amélioration dans la situation, la position de qqn.** → **Avancement. L'amélioration des relations entre personnes, entre pays.** → **Détente.** *Il s'est produit une amélioration dans les relations de ces deux pays. Apporter, faire, opérer, réaliser une amélioration.* — *L'amélioration morale, spirituelle, intellectuelle de...*

Je tâcherais de deviner la distance où nous sommes encore de l'accomplissement total de l'Évangile, en supputant le nombre des maux détruits et des améliorations opérées dans les dix-huit siècles écoulés (...).

CHATEAUBRIAND, *Mémoires d'outre-tombe*, t. II, p. 209.

Cette impression de progrès inspirait l'optimisme et justifiait l'espoir d'une amélioration plus rapide.

Ch. SEIGNOBOS, *Hist. sincère de la nation franç.*, p. 234.

(...) inventions humaines qui tendent à l'amélioration de la vie individuelle et sociale.

G. DUHAMEL, *Scènes de la vie future*, XII, p. 198.

L'amélioration par la souffrance est une supposition chrétienne et je ne suis pas chrétien.

Ch. PÉGUY, *Œ.*, t. XI, p. 106 à 108.

♦ 2 Agric. Action d'améliorer (un sol). → **Améliorant; abonissement, amendement, bonification.**

♦ 3 (1510). Dr. (au plur.). Ensemble de travaux ou de dépenses faits sur un bien et lui procurant une plus-value. → **Plus-value. Améliorations faites sur un immeuble par une personne tenue de le restituer.** → **Impense. Améliorations nécessaires, utiles, voluptueuses. Faire des améliorations dans une maison, un château, un jardin...** → **Embellissement, réparation, restauration.**

(...) toutes les réparations et améliorations qu'il aura faites au fond.

(...) toutes les améliorations survenues à l'immeuble hypothéqué.

Code civil, art. 1634.
Code civil, art. 2133.
CONTR. *Adulteration, aggravation, avarie, corruption, déclin, dégénération, dégénérescence, dégradation, déprava- tion, détérioration, endommagement, perversion, pervertissement.*

Annexe 5 (Chap. 5) : articles lexicographiques de performance.

Article lexicographique extrait du Larousse Lexis :

PERFORMANCE [pɛrformɑ̃s] n. f. (mot angl., de l'anc. fr. *parformance*, de *parformer*, accomplir; 1839). 1. Résultat obtenu par un athlète, une équipe dans une épreuve sportive ou dans un match, par un cheval dans une course : *Réaliser une belle performance. On attend de notre équipe du 4 × 100 une performance. La performance ne sera pas homologuée.* — 2. Résultat obtenu au cours d'une épreuve, d'un test. — 3. (v. 1960). Ling. Manifestation de la compétence des sujets parlants dans leurs multiples actes de parole. — 4. Exploit : *J'avais voué une haine spéciale aux spéléologues, dont les performances m'écœuraient* (Camus). — 5. Test de performance, épreuve non verbale destinée à apprécier les fonctions intellectuelles. ♦ n. f. pl. 1. Aéron. Chiffres et courbes qui caractérisent les possibilités de vol d'un avion. — 2. Qualités (accélération, vitesse maximale, décélération) qui caractérisent la valeur d'une voiture. ♦ **performant**, e adj. (v. 1960). 1. Se dit d'un système, d'un appareil, d'une machine susceptible d'un rendement particulièrement élevé. — 2. Econ. Se dit d'une entreprise, d'un produit très compétitif sur le marché. ♦ **contre-performance** n. f. [Pl. 2.] En sport, et, partic., au tennis, défaite subie par un joueur classé devant un joueur non classé ou devant un joueur moins bien classé que lui.

Article lexicographique extrait du Grand Robert :

PERFORMANCE [pɛʁfɔʁmãs] n. f. — 1839; mot angl., «accomplissement»; anc. franç. *parformance* (xvi^e), même sens, de *parformer* (déb. xiii^e) «accomplir, exécuter; achever».

♦ 1 [a] Turf. Manière dont se comporte un cheval de course au cours d'une épreuve. *Bonne, mauvaise performance.* — Spécialit. Résultat chiffré obtenu au cours de cette épreuve. (Au plur.) Tableau des épreuves subies par le même cheval. — REM. Le mot n'est enregistré qu'au pluriel dans Littré.

[b] (1876). En parlant des athlètes, des cyclistes, etc., au cours d'une épreuve sportive. *Accomplir, réaliser une performance. Performance homologuée* (cit. 3). *Joueur de tennis classé d'après ses performances. Les performances d'un champion. C'est une médiocre performance pour un cycliste de sa classe.* — (xx^e). Par ext. *Performance d'une automobile. Voiture classée première à l'indice de performance, relativement à sa cylindrée.*

Les Grecs (...) n'ont gravé sur le piédestal d'aucune de leurs statues d'athlètes les performances du bel humain glorifié. MONTHERLANT, *les Olympiques*, p. 57.

(...) ayant piqué un galop au sortir du village et semé tous ses officiers, il franchit une haie d'un bond (...) précédant de dix longueurs sa suite mortifiée. Le capitaine-adjoint ne fut pourtant pas ébloui par cette performance.

R. DORGELES, *le Cabaret de la belle femme*, p. 161. Sans doute son désir était-il peu fondé, puisque, si elle battait ce record, sa performance, accomplie sans témoins officiels, ne serait pas homologuée.

MONTHERLANT, *les Olympiques*, éd. L. de Poche, p. 77.

(1869). Par anal. Manière de développer un sujet, d'exécuter une œuvre, en public.

C'est au fond à la «performance» que j'en ai. Ce Leonard Bernstein en fait trop, pour mon goût, mais c'est parce qu'il n'y a rien qu'il ne puisse faire.

F. MAURIAC, *le Nouveau Bloc-notes 1958-1960*, p. 128.

♦ 2 (1924). Résultat sportif exceptionnel. → *Record.* — (V. 1928). Exploit, réussite remarquable dans tout autre domaine. *Le travail a été exécuté en moins de temps qu'il n'était prévu, c'est une belle performance! Performances amoureuses, érotiques.*

♦ 3 Psychol. (Anglic.). Résultat individuel dans l'accomplissement d'une tâche, dont les facteurs principaux sont l'aptitude et la motivation. *Besoin de performance de l'individu. Niveau de performance* : degré de réussite individuelle. *Index de performance*, évaluant le besoin de performance et permettant un jugement sur l'ambition, la persévérance de quelqu'un. — *Test de performance* : épreuve d'intelligence non verbale.

♦ 4 (1943. In Höfler). Techn. Rendement maximal (d'une machine, et, par ext., d'un être vivant, de l'homme). *Performance d'un système électronique, d'un ordinateur.* → *Performant.* — Spécialit. Indications quantitatives sur les caractéristiques d'un véhicule, d'un système mécanique... *Les performances d'un avion* (vitesse de croisière, de montée maximale; autonomie; rayon d'action, etc.), *d'une automobile* (accélération, vitesse d'accélération), *d'une machine-outil* (nombre de tours/minute, vitesse de coupe, etc.).

Écon. Rendement maximal; capacités de production (d'une entreprise, d'une méthode de travail, etc.).

♦ 5 (De l'angl.; mot employé dans ce sens par le linguiste américain N. Chomsky). Ling. Réalisation actuelle, sous forme d'échange d'informations (de messages), des virtualités du langage (appelées dans cette terminologie : *compétence*). *Le modèle linguistique de performance étudie la production des phrases réellement observables, avec tous leurs caractères concrets* (variations dialectales et individuelles, erreurs, lapsus, limitations en durée) *alors que la théorie grammaticale rend compte de la compétence.*

DÉR. *Performant.*

PERFORMANT, ANTE [pɛʁfɔʁmã, ãt] adj. — 1968, Larousse; de *perform(ance)*.

Techn. (anglic.). Dont le niveau de performances est élevé, qui est riche de possibilités, en parlant d'un système technique (d'abord, électronique). *Ordinateur individuel particulièrement performant.* «Si vous choisissez une calculatrice aux capacités trop réduites, vous serez conduit (...) à vous en procurer une autre plus performante» (Sciences et Avenir, n° 35, p. 32). Par ext. *Une entreprise performante*, (Qualifiant une personne). *Société en pleine expansion recherche pour son réseau étranger directeur des ventes agressif* et performant.*

Annexe 6 (Chap. 5) : articles lexicographiques de erreur.

Article lexicographique extrait du Larousse Lexis :

ERREUR [ɛʁœʁ] n. f. (lat. *error*, de *errare* [v. **ERRER**]; 1130).
1. Action de se tromper; faute commise en se trompant : *L'erreur est la règle; la vérité est l'accident de l'erreur* (Duhamel). *C'est tout ce que vous leur reprochez? une erreur de calcul?* (Beauvoir). *Le candidat a commis une erreur importante en situant Poitiers sur la Loire* (syn. **BÉVUE**, **MÉPRISE**). *Rectifier, relever une erreur* (syn. **INEXACTITUDE**). *Cette bataille a eu lieu, sauf erreur, en 1745* (= si je ne me trompe pas). *Par suite d'une erreur matérielle, ce paragraphe a été omis* (= une erreur qui ne procède pas d'un raisonnement faux; syn. **INADVERTANCE**). — **2.** (1170). État de quelqu'un qui se trompe : *Son erreur confirmerait, s'il en était besoin, la seule victoire que nous ayons remportée depuis un demi-siècle* (Mauriac). *Vous êtes dans l'erreur la plus complète* (= votre opinion est totalement fausse). *Il persiste dans son erreur. Une information incomplète m'avait induit en erreur.* — **3.** (1480). Action faite mal à propos, inconsidérée, regrettable : *Cette intervention fut une erreur* (syn. **GAFFE**, **MALADRESSE**). *L'erreur a été de la laisser trop longtemps dans ses vêtements trempés* (Gide) [syn. **↑ABERRATION**]. *Un écrivain qui considère son premier roman comme une erreur de jeunesse* (syn. **ÉGAREMENT**). — **4.** (1864). *Métrol.* Différence entre la valeur exacte d'une grandeur et la valeur donnée par la mesure. — **5.** (1690). *Dr.* Vice du consentement qui rend le contrat annulable lorsqu'il porte sur les qualités substantielles. — **6.** *Phys.* *Erreur absolue*, différence entre un nombre et sa valeur approchée, entre le résultat de la mesure d'une grandeur et la valeur exacte de cette grandeur. || *Erreur judiciaire*, condamnation prononcée à tort contre un innocent par suite d'une erreur de fait, et qui ouvre le pourvoi en révision. || *Erreur n'est pas compte*, formule par laquelle une personne qui s'est trompée dans un compte rappelle qu'il n'y avait là rien d'intentionnel. || *Erreur relative*, rapport de l'erreur absolue à la valeur de la grandeur elle-même. || *Erreur des sens*, illusion. || *Phys.* *Erreur systématique*, erreur provenant d'une cause permanente (méthode employée, mauvais réglage d'un instrument, etc.). • **LOC. ADV.** *Par erreur*, en se trompant, par ignorance ou par étourderie : *Nous avons pris par erreur la route qui menait à la rivière* (syn. plus recherché **PAR MÉGARDE**). *Il est entré dans cette salle par erreur* (= sans le faire exprès). ♦ n. f. pl. (1355). Actes condamnables, dérèglements : *Il a bien des erreurs à se faire pardonner.* ♦ **erroné**, **e** adj. (lat. *erroneus*; v. 1400). Qui comporte une erreur : *Qu'elle soit [cette idée], en soi, insuffisante, et, dans une certaine mesure, erronée, d'accord* (Adamov) [syn. **FAUX**; contr. **EXACT**]. *Un calcul erroné* (syn. **FAUTIF**, **INEXACT**; contr. **IUSTE**). ♦ **erronément** adv. (v. 1500). Rare. *Déclarer erronément*.
 • **CLASS. erreur** n. f. **1.** Randonnée, vaste voyage comportant des aventures : *Les erreurs d'Ulysse* (Académie). — **2.** Illusion : *Il faut que l'erreur de l'avenir nous soutienne* (Massillon). — **3.** *Mettre hors d'erreur*, détromper : *Et je veux bien tous deux vous mettre hors d'erreur* (Molière).

Article lexicographique extrait du Grand Robert :

ERREUR [ɛʁœʁ] n. f. — V. 1125, *error*, fin x^e, lat. *error*, de *errare*, fig. → *Erre*.

❶ 1 (Une, des erreurs; l'erreur). Acte de l'esprit qui tient pour vrai ce qui est faux* et inversement; jugements, attitudes psychiques qui en résultent. → *Errement*, *faute*, *fourvoiement*. *Erreur choquante*, *grossière*, *commise par ignorance*. → *Anerie*, *bêtise*, *bourde*. *Erreur par laquelle on confond plusieurs choses, plusieurs personnes*. → *Bévue*, *confusion*, *malentendu*, *méprise*, *quiproquo*. *La Comédie des erreurs*, pièce de Shakespeare. *Erreur capitale*, *complète*, *totale*. *Profonde erreur* (→ *Borne*, cit. 5), *grosse erreur* (→ *Conquérir*, cit. 12). *Erreur extrême*, *pernicieuse* (→ *Attendre*, cit. 92 et 107). *Erreur volontaire*. *L'homme est sujet à l'erreur*, à se tromper (→ *Apprécier*, cit. 4). → *Errare humanum est*. *Faire, commettre une erreur*. → *Abuser* (s), *égarer* (s), *errer* (l.), *faillir*, *méprendre* (se), *tromper* (se). *Revenir d'une erreur*. → *Désabuser* (se). *Éviter une erreur*, *éviter l'erreur en ne portant pas de jugement* (→ *Caractère*, cit. 20). *Erreur d'appréciation*, *d'estimation* (→ *Mésestimer*), *de jugement* (→ *Aberration*; *méjuger*), *de raisonnement*. → *Paralogisme*, *sophisme*, *vici* (de raisonnement). *Erreur de goût*, *contre le goût* (→ *Défaut*, *faute*, *manquement*). *Erreur de prononciation* (→ *Cuir*), *erreur qui fait dire un mot pour un autre* (→ *Lapsus*). — *C'est une erreur de croire, que de croire cela*. *Vous croyez cela* : grave erreur, *profonde erreur*. — *Convaincre qqn d'erreur*, lui prouver qu'il s'est trompé.

C'est erreur, ou plutôt c'est crime de le croire. LA FONTAINE, *Fables*, II, 13.

La grande erreur des gens d'esprit est de ne pas croire le monde aussi bête qu'il est.

M^{me} DE TENCIN, citée par CHAMFORT, *Caractères et anecdotes*.

Les erreurs de nos jugements produisent l'ardeur de tous nos désirs. ROUSSEAU, *Émile*, IV.

C'est une grande erreur de spéculer sur la sottise des sots (...). VALÉRY, *Rhumbs*, p. 244 (→ *Bâtir*, cit. 36).

C'était une idée à laquelle il tenait : qu'il y a, fatalement, à la base de tout amour passionné, un malentendu, une illusion généreuse, une erreur de jugement ; une conception fautive qu'on s'est faite l'un de l'autre et sans laquelle il ne serait pas possible de s'aimer aveuglément.

MARTIN DU GARD, *les Thibault*, t. IX, p. 45.

Nous commettons sans cesse, dans l'ordre de la pensée, des erreurs analogues à celle qui nous ferait prendre, dans l'ordre des sens, un homme normal pour un nain, parce que nous le voyons de loin.

PAULHAN, *Entretien sur des faits divers*, p. 99. *Errare humanum est*. L'erreur est essentielle à l'humanité, c'est un élément de progrès et de découverte ; qui élimine l'erreur élimine la recherche, cette provocation à l'inconnu et l'inattendu, quelque chose d'étranger à qui nous fournissons le moyen d'intervenir.

CLAUDEL, *Journal*, févr.-mars 1930.

loc. FAIRE ERREUR : se tromper. Vous faites erreur.

— IL Y A ERREUR. Il y a erreur sur la personne (→ *Idem*, Gourance). *Fam. Il n'y a pas (y a pas) d'erreur* : c'est bien cela. — SAUF ERREUR : excepté si l'on se trompe (→ *infra* II, 2). *Sauf erreur de ma (ta, sa, votre...) part*.

Erreurs des sens : erreurs d'interprétation des données des sens. → *Illusion*; *apparence* (fausse). → *Courrier*, cit. 2.

Quant à l'erreur et incertitude de l'opération des sens, chacun s'en peut fournir autant d'exemples qu'il lui plaira, tant les fautes et tromperies qu'ils nous font sont ordinaires. MONTAIGNE, *Essais*, II, XII.

Je ne veux plus me retenir des erreurs de mes doigts, des erreurs de mes yeux. Je sais maintenant qu'elles ne sont pas que des pièges grossiers, mais de curieux chemins vers un but que rien ne peut me révéler, qu'elles. A toute erreur des sens correspondent d'étranges fleurs de la raison.

ARAGON, *le Paysan de Paris*, p. 13.

Fig. *Les erreurs du cœur*. *Les Erreurs amoureuses*, recueil de poèmes de Pontus de Thiard.

Le cœur se trompe, comme l'esprit ; ses erreurs ne sont pas moins funestes et l'on a plus de mal à s'en défaire à cause de la douleur qui s'y mêle.

FRANCE, *le Petit Pierre*, V, p. 26.

PAR ERREUR. *Faire qqch., dire qqch. par erreur*. → *Mégarde* (par). *Il a été condamné par erreur*, à la suite d'une erreur (II) judiciaire.

❷ 2 (l'erreur : surtout dans des constructions avec *dans*, *en*, *de*...). État d'un esprit qui prend pour vrai ce qui est faux, et inversement (soit qu'il se trompe soit qu'il soit trompé). → *Aberration*, *aveuglement* (→ *Aveugler*, cit. 21). *Être dans l'erreur*. *Donner, tomber dans l'erreur*. *Être embourbé, enfoncé dans l'erreur*. *Rester, persévérer dans l'erreur par préjugé, prévention, idée préconçue*. *Pousser* (cit. 25) *l'erreur jusqu'au bout*. *Induire en erreur*. → *Blouser*, *fourvoyer*, *séduire*, *tromper*. *Laisser dans l'erreur*. *Tirer qqn d'une erreur*. → *Détromper* (→ *Asses*, cit. 39). *Sortir de l'erreur pour entrer dans le doute*.

(...) l'erreur n'est pas une pure négation, c'est-à-dire n'est pas le simple défaut ou manquement de quelque perfection qui ne m'est point due, mais c'est une privation de quelque connaissance qu'il me semble que je devrais avoir. DESCARTES, *Méditations*, IV, 4.

C'est trop vous tenir dans l'erreur ; il faut lever le masque (...). MOLIERE, *la Princesse d'Elide*, V, 2.

Induire en erreur est mettre l'homme dans la nécessité de conclure et suivre une fausseté.

PASCAL, *Pensées*, XIII, 821.

L'erreur est la nuit des esprits et le piège de l'innocence. VAUVENARGUES, *Réflexions et maximes*, 388.

(...) les ténèbres de l'ignorance valent mieux que la fausse lumière de l'erreur (...).

ROUSSEAU, *Lettre à Mgr de Beaumont*.

L'erreur agite ; la vérité repose.

Joseph JOUBERT, *Pensées*, XI, LXXIII.

Mieux vaut l'erreur que le doute, — pourvu qu'elle soit de bonne foi.

R. ROLLAND, *Musiciens d'aujourd'hui*, p. 118.

♦ 3 (L'erreur; une erreur). Ce qui, dans ce qui est perçu ou transmis comme étant vrai (apparences, connaissances), est jugé comme faux par celui qui parle. → **Fausseté, illusion, mensonge, songe.** Les **attraits** (cit. 7) de l'erreur. Donner à l'erreur une apparence flatteuse, l'apparence de la vérité. → **Sophisme** (→ Colorer, cit. 9). Auteurs qui répandent le mensonge et l'erreur (→ **Charlatan**, cit. 4). Toute cette théorie n'est qu'erreur; est entachée d'erreur (→ **Erroné, faux**). Tout n'est qu'erreur et mensonge. → **Trompeur, vain, vanité.** C'est une erreur pour moi : je juge cela faux.

(...) l'homme (...) dépositaire du vrai, cloaque d'incertitude et d'erreur; gloire et rebut de l'univers.

PASCAL, *Pensées*, VII, 434 (→ **Chaos**, cit. 4).

(...) la grandeur est un songe, la joie une erreur (...)

BOSSUET, *Oraison funèbre de Marie-Thérèse d'Autriche* (→ **Trompeur**).

Ce monde est-il le vrai? le nôtre est-il l'erreur? O possibles qui sont pour nous les impossibles!

HUGO, *les Contemplations*, III, xxx, II. Ce qu'un système affirme, c'est sa part de vérité; ce qu'il nie, c'est sa part d'erreur.

RENAN, *L'avenir de la science*, in *Œ. compl.*, t. III, p. 1085.

(...) ceux (les intellectuels) dont le métier est de chercher la vérité au milieu de l'erreur (...)

R. ROLLAND, *Au-dessus de la mêlée*, p. 12.

(...) que la vérité est l'un soit l'erreur de l'autre, quelle meilleure conclusion? « Qu'est-ce que la vérité? » dit Ponce-Pilate.

André SUARES, *Trois hommes*, « Ibsen », VIII, p. 169.

Nul doute : l'erreur est la règle : la vérité est l'accident de l'erreur.

G. DUHAMEL, *Chronique des Pasquier*, I, p. 26.

L'erreur s'accompagne de certitude. L'erreur s'impose par l'évidence. Et tout ce qui se dit de la vérité, qu'on le dise de l'erreur : on ne se trompera pas davantage. Il n'y aurait pas d'erreur sans le sentiment même de l'évidence. Sans lui on ne s'arrêterait jamais à l'erreur.

ARAGON, *le Paysan de Paris*, p. 9.

Absolt. L'erreur : ce qui est erroné (par rapport à, pour qqn).

Vérité au delà des Pyrénées, erreur au delà.

PASCAL, *Pensées*, VI, 294 (→ **Degré**, cit. 9).

♦ 4 (Une, des erreurs). Assertion, opinion fautive (→ **Contre-vérité, fausseté**). Erreur commune, courante, fréquente; erreur populaire. Erreur ancienne, vieille comme le monde. Erreur répandue, enracinée (→ **Comment**, cit. 12). **Avancer** (cit. 9); **soutenir** une erreur. Aimer, chérir, diviniser ses erreurs. Combattre des erreurs. Reconnaître, confesser, corriger, rectifier ses erreurs (→ **Plebéien**, cit. 7). Il est revenu de bien des erreurs. → **Préjugé**. Manuel, *compendium* (cit. 2) d'erreurs (→ **Absurdité**, cit. 3).

Il n'y a point de prescription contre la vérité : les erreurs pour être vieilles n'en sont pas meilleures.

BAYLE, *Nouvelles de la République des Lettres*, mars 1684, Deux dissertations sur oracles païens.

Il y a des esprits qui vont à l'erreur par toutes les vérités; il en est de plus heureux qui vont aux grandes vérités par toutes les erreurs.

Joseph JOUBERT, *Pensées*, XI, LXVIII.

(...) propager une erreur qu'on sait mortelle pour la Société.

J. ROMAINS, *les Hommes de bonne volonté*, t. III, I, p. 22.

♦ 5 **Absolt. L'erreur**. Conviction, doctrine qui s'écarte d'un dogme, au regard de ceux qui le défendent. Erreur en matière de convictions religieuses, de foi (→ **Hérésie**). Tomber, vivre dans l'erreur, persister dans l'erreur (→ **Impénitence**). **Prédicateur qui combat l'erreur, triomphe de l'erreur. Renier l'erreur** (→ **Abjurer**, cit. 2). **Faire régner l'erreur** (→ **Arracher**, cit. 10). **Défendre** (cit. 8) l'erreur.

L'erreur et la nouveauté se faisaient entendre dans toutes les chaires (...)

BOSSUET, *Oraison funèbre de la reine d'Angleterre*.

Ce Dieu, maître absolu de la terre et des cieux, N'est point tel que l'erreur le figure à vos yeux.

RACINE, *Esther*, III, 4.

Pendant que tous les philosophes se séparent en différentes sectes, il se trouve en un coin du monde des gens qui sont les plus anciens du monde, déclarant que tout le monde est dans l'erreur, que Dieu leur a révélé la vérité, quelle sera toujours sur la terre.

PASCAL, *Pensées*, IX, 618.

Si les hérétiques couverts, et qui, ne se séparant pas de l'Eglise, sont néanmoins dans l'erreur (...)

PASCAL, *Pensées*, XIII, VIII.

♦ 6 (Une, des erreurs). Action regrettable, maladroite, déraisonnable. → **Faute**. Il a commis une grossière erreur en négligeant de l'inviter, en se trompant d'adresse. → **Etourderie, gaffe, impair, inadvertance, inattention, maladresse, manquement, oubli. Erreur d'étiquette, de savoir-vivre.** — **Erreur de tactique.** — **Spécialt.** Écart de conduite; action blâmable (et jugée comme telle par celui qui l'a commise). → **Dérèglement, écart, égarement, errements** (cit. 3), **extravagance. Erreurs de jeunesse. Son mariage est une erreur. Subir les conséquences d'une erreur passée. Les erreurs d'une âme faible.** → **Défaillance, faiblesse. L'erreur que l'amour, la passion lui a fait faire. Pardonner les erreurs passées.**

(...) un amour que je regardais plutôt comme une erreur de l'enfance (...)

LACLOS, *les Liaisons dangereuses*, Lettre XCVIII.

(...) réparer les erreurs de sa jeunesse (...)

LACLOS, *les Liaisons dangereuses*, Lettre CXXIV.

Dites-vous bien que si la Légion est un refuge, c'est à la condition de racheter les erreurs du passé (...) pour ceux qui, naturellement, ont un passé chargé d'erreurs.

P. MAC ORLAN, *la Bandera*, VIII, p. 96.

Spécialt. Faute, péché. **Se confesser de ses erreurs.**

II (Sens objectif). **A** Une, des erreurs. ♦ **I** Chose fautive, erronée par rapport à une norme (différence par rapport au modèle ou au réel). Faute, inexactitude. Relever une erreur dans un texte, un passage, une citation, une référence. Corriger, effacer, rectifier, redresser une erreur. Correction des erreurs (→ Avis, cit. 31). Erreur typographique, erreur d'impression (→ Bourdon, coquille, doublon, mastic). Liste d'erreurs (→ Errata). Erreur dans une traduction, une interprétation. → Contresens, faux-sens, non-sens. Erreur de date, de temps. → Anachronisme (cit. 1), métachronisme. Texte rempli, bourré d'erreurs, contenant de nombreuses erreurs. → Fautif. Cette erreur est due au défaut d'attention, à la négligence, à l'inadvertance du correcteur. Faire une erreur en composant un numéro de téléphone. Raccrochez, c'est une erreur!

Renduel veut en publier une deuxième édition complète. Brizeux avait remarqué, à la première édition complète qui fut faite très négligemment, des erreurs et inadvertances impardonnables (...).

SAINT-BEUVE, Correspondance, 454, 7 mars 1835.

♦ **2** Chose erronée, élément inexact, dans certaines opérations particulières. Erreur dans un compte. → Mécompte. Rectifier une erreur par une écriture rectificative. Sauf erreur ou omission : formule que l'on met au bas des comptes courants et qui réserve le droit du possesseur du compte à le vérifier. — Prov. Erreur n'est pas compte*. — Erreur volontaire. → Falsification, irrégularité. — Erreur dans la distribution des cartes. → Maldonne.

Erreur de calcul : inexactitude dans un calcul.

L'erreur de calcul dans une transaction doit être réparée. Code civil, art. 2058.

Phys. Écart entre la valeur exacte d'une grandeur et sa valeur calculée ou mesurée. Erreur systématique, due à la méthode ou à l'instrument utilisé; erreur accidentelle. Erreur absolue : la quantité dont il faut augmenter ou diminuer la valeur de la mesure pour que le nombre réel soit compris entre la valeur de la mesure augmentée de l'erreur et cette même valeur diminuée de l'erreur. Erreur relative : rapport de l'erreur absolue à la valeur de la mesure. Rapport de la valeur de la mesure à l'erreur absolue : degré d'approximation de la mesure. L'erreur absolue commise sur une somme, une différence est égale à la somme des erreurs absolues commises sur chacun des termes. L'erreur relative commise sur un produit, un quotient est égale à la somme des erreurs relatives commises sur chacun des termes. — Courbe d'erreur. — Expérience entachée d'erreurs.

(...) il n'y a aucune circonstance (...) dans laquelle on puisse affirmer que des êtres microscopiques sont venus au monde sans germes (...) Ceux qui le prétendent ont été le jouet d'illusions, d'expériences mal faites, entachées d'erreurs qu'ils n'ont pas su percevoir ou qu'ils n'ont pas su éviter. PASTEUR, in Henri MONDOR, Pasteur, p. 71.

Loi des erreurs.

La loi des erreurs admise par tous les calculateurs est la loi de Gauss, qui est représentée par une certaine courbe transcendante connue sous le nom de « courbe en cloche ». Henri POINCARÉ, la Science et l'Hypothèse, p. 240.

Psychol. Erreurs individuelles d'observation. → Équation (équation personnelle).

Dr. → Vice. Erreur portant sur un élément du contrat. Erreur sur la nature du contrat, erreur sur l'objet. Erreur sur la substance; erreur sur la personne. Erreur de fait, portant sur une circonstance matérielle (→ Amen, cit. 28). Erreur de droit, portant sur l'existence ou l'interprétation d'une règle juridique. En droit pénal, l'erreur de droit n'a pas de conséquence (cf. Nul n'est censé ignorer la loi). Erreur sur la substance, qui porte sur les qualités déterminantes de l'objet du contrat.

Lorsqu'il y a eu erreur dans la personne, le mariage ne peut être attaqué que par celui des deux époux qui a été induit en erreur. Code civil, art. 180.

Il n'y a point de consentement valable, si le consentement n'a été donné que par erreur (...) L'erreur n'est une cause de nullité de la convention que lorsqu'elle tombe sur la substance même de la chose qui en est l'objet. Elle n'est point une cause de nullité, lorsqu'elle ne tombe que sur la personne avec laquelle on a l'intention de contracter, à moins que la considération de cette personne ne soit la cause principale de la convention.

Code civil, art. 1109-1110.

Cour. Erreur judiciaire : erreur de fait commise par le juge et entraînant la condamnation d'un innocent. — (Il y a) erreur sur la personne ou, absolu, il y a erreur.

B L'erreur : l'ensemble des erreurs. Sc. Courbe d'erreur. → ci-dessus, 2. Dr. L'erreur, vice de consentement. Nullité d'un acte entaché d'erreur.

III (XVI^e). Vx. Action d'errer ça et là. → Errance, pérégrination.

Encore que nous disions des Étoiles errantes, nous ne disons pourtant point l'erreur des Étoiles (...) Erreur est proprement en notre langue une fausse opinion (...) nous sommes plus retenus que ceux qui ont dit les erreurs d'Ulysse.

CHEVREAU, Observations sur Malherbe (1660), I, 276.

Comtez-moi d'Illion les terribles assauts,
Et vos longues erreurs sur la terre et sur l'onde.

DELILLE, Énéide, I.

Il se pourrait que vous m'eussiez écrit, car, dans mes longues erreurs, j'ai perdu des lettres.

P.-L. COURIER, Lettres, I, 322.

Fig. et vx. Erreur de l'imagination, du cœur (compris aujourd'hui au sens I.).

CONTR. Justesse, lucidité, perspicacité. — Certitude, exactitude, réalité, vérité. — Correction, rectification.

Annexe 7 (Chap. 5) : articles lexicographiques de problème.

Article lexicographique extrait du Larousse Lexis :

PROBLÈME [prɔblɛm] n. m. (lat. *problema*; v. 1380). 1. Dans le domaine scientifique, question qui appelle une solution d'ordre logique, rationnel : *Un problème mathématique. Les termes du problème. Les données d'un problème. Faire un problème de géométrie* (syn. DEVOIR). — 2. (1753). Tout ce qui est difficile à résoudre, à expliquer : *Le problème moral* (= les questions que se pose l'homme sur la morale). *Pour lui, le Mal n'était pas un problème, mais un mystère* (Malraux). *Cet homme est un problème* (= on ne peut comprendre sa conduite). *Le problème technique ne s'y pose pas* (Mauriac). *Les problèmes de la circulation dans Paris. Le problème lui parut insoluble et probablement dénué d'intérêt* (Aymé). *Chacun a ses problèmes* (syn. DIFFICULTÉ, ↑ENNUI). — 3. Fam. *Il n'y a pas de problème*, il ne faut pas hésiter, c'est très simple. ◆ **problématique** n. f. (v. 1900). 1. Ensemble des questions posées par une branche de la connaissance : *La problématique de ce philosophe est au fond très simple.* — 2. Philos. Possibilité théorique d'un problème philosophique. ◆ adj. *Jugement problématique*, dans la philosophie de Kant, jugement dans lequel le rapport de l'attribut au sujet n'est conçu que comme simplement possible. ◆ **problématisation** n. f. Philos. Opération par laquelle un philosophe transforme l'expérience personnelle qu'il a d'un phénomène humain général en un énoncé visant à remettre en question les bases présumées de ce phénomène.

Article lexicographique extrait du Grand Robert :

PROBLÈME [pʁɔblɛm] n. m. — 1382; lat. *problema*, grec *problēma*.

♦ 1 Question à résoudre, point obscur que l'on se propose d'éclaircir, qui prête à discussion, dans un domaine quelconque de la connaissance.

→ **Question**. *Problèmes philosophiques, moraux* (→ Distinguer, cit. 2), *métaphysiques* (1. Métaphysique, cit. 3). *Le problème du mal* (3. Mal, cit. 50), *de l'origine* (cit. 10) *des idées*. *Soulever le problème de...* Il a très bien vu le problème. — C'est là la clef du problème. Le problème est clair, est bien posé. *Problème verbal*. *Problème insoluble, soluble*. *Aborder* (→ 1. Avocat, cit. 17), *traiter*, *résoudre un problème*. — *Le vieux problème de l'origine* (cit. 11) *des religions* loc. *Tout le problème est là* : c'est toute la question.

Nous espérons y trouver quelques données pour la solution de ce problème historique : comment le protestantisme s'est infiltré dans les grandes maisons de France, et comment s'est formée cette double opposition populaire et aristocratique (..).

BALZAC, le Feuilleton, in OE. diverses, t. I, XXXV, p. 419.

Je confesse que j'ai coutume de distinguer dans les problèmes de l'esprit ceux que j'aurais inventés et qui expriment un besoin réellement ressenti par ma pensée, et les autres, qui sont les problèmes d'autrui.

VALÉRY, Variété, in OE., Pl., t. I, p. 1318.

«TOUT LE PROBLÈME EST LÀ» Brevet de bon sens, de parfaite logique, décerné au parleur qui vient de terminer sa démonstration politique, sociale ou économique. Ex. : «Le Marché Commun, c'est très fol, encore faut-il que l'on joue le jeu franchement ! — Tout le problème est là...»

Pierre DANINOS, Un certain Monsieur Blot, p. 237.

Spécialt (dans les sciences mathématiques). Question à résoudre, portant soit sur un résultat inconnu à trouver à partir de certaines données, soit sur la détermination de la méthode à suivre pour obtenir un résultat supposé connu. *Problèmes mathématiques* (→ Équation, cit. 2). *Problèmes de géométrie, d'algèbre* (→ Opération, cit. 3), *de mécanique* (cit. 8) *céleste*. *Dégager l'inconnue d'un problème algébrique*. *Déduire* (cit. 5) *la solution de la supposition du problème résolu*. → **Analyse** (infra cit. 8). *Résoudre un problème* (→ Approximation, cit. 1). *Résolution d'un problème par la méthode de...* *Problème qui comporte deux, trois... solutions*. *Problème indéterminé*. *Problème insoluble*. → **Aporisme**. *Problème de la quadrature du cercle*. *Problème des partis* (traité par Pascal et Fermat). *Problème (ou paradoxe) de Saint-Petersbourg*. *Le problème du tricheur* (dû à Henri Poincaré).

REM. Un grand nombre de problèmes mathématiques sont désignés par le nom des savants qui les ont proposés ou résolus. *Problème de Pappus, problème de Cauchy*.

(1913). Exercice scolaire comportant des questions à résoudre. *Problème d'arithmétique* (→ Calcul), *de mathématiques* (→ Classe, cit. 13). *Les données*, l'énoncé** d'un problème. *Questions de cours et problèmes*. *Faire un problème*. *Tu as fait, fini tes problèmes?* *Problème de robinets** (cit. 1, 2, 1, 3 et supra).

On nous avait donné — pour modérer nos plaisirs de vacances — quelques problèmes à secret, dont la solution, je le savais bien, résultait toujours d'une astuce tapie sournoisement dans l'énoncé. Ce dernier présentait un visage benin et ironique, comme une devinette. Il avait pour mission de vous fourvoyer, en désespoir de cause, dans des opérations interminables dont le résultat (s'il en résultait, par hasard, quoi que ce fût) était monumentalement absurde. Et le maître vous le démontrait sèchement au tableau.

H. BOSCO, Antonin, p. 89.

(Dans les jeux dont les règles sont précises). *Problème d'échecs*. *Cette revue, ce journal propose des problèmes de mots croisés, de bridge*.

Faux problème : problème mal posé, posé à côté du vrai problème, qui ne correspond pas aux vraies difficultés (→ 1. Faux, cit. 31).

Le danger, aujourd'hui, c'est qu'on se pose de fausses questions. On résout des milliers de problèmes. Mais, hélas ! ce sont de faux problèmes dès le départ.

J.-L. GODARD, Arts, n° 716, 1^{er} AVR. 1959, in Coll. des Cahiers du cinéma, p. 237.

♦ 2 (Dans l'ordre de la pratique, de l'action). Difficulté qu'il faut résoudre pour obtenir un résultat ; situation instable ou dangereuse exigeant une décision.

→ **Ennui**. *Problèmes techniques* (→ Culture, cit. 2). *Les problèmes de la circulation* (cit. 5), *du stationnement*. *C'est un problème de digestion, de foie...* *C'est un simple, un petit problème de trésorerie*. → (Art, littér.). *Problèmes plastiques* (→ Expressionnisme, cit. 2).

(...) unité dans l'aspect et variété infinie dans le détail, voilà le difficile problème que les artistes du moyen âge ont presque toujours résolu avec bonheur.

Th. GAUTIER, Voyage en Espagne, p. 27.

Être débordé par les problèmes pratiques (→ Inopiné, cit. 2). *Régler un problème* : résoudre une difficulté d'ordre pratique.

loc. *Faire problème* : présenter des difficultés. *Son refus ferait problème*. — *Poser des problèmes* : susciter des difficultés qu'il faudra résoudre. → **Bémol** (3.), hic. *Cet enfant me pose des problèmes*. *Ça pose, ça me pose des problèmes !* — (Récrit, prétentieux). *Ça pose problème*.

(Calque de l'angl. des États-Unis; av. 1959). *Fam. Il n'y a pas, (fam.) y a pas [japa] de problème* : c'est une chose simple, évidente ou facile; s'emploie aussi en réponse : bien sûr, évidemment. «Tu pourras venir demain ? — Pas de problème !», certainement.

C'est mon (ton, son) problème : cela me (te, le) concerne particulièrement.

— Ça, coco, c'est ton problème.

— On dit toujours ça pour se désintéresser.

E. AJAR (R. GARY), l'Angoisse du roi Salomon, 1979, p. 84.

Spécialt. [a] (Appliqué à une collectivité). *Problèmes politiques* (→ Nationaliste, cit. 3). *Le problème de la Tchécoslovaquie* (→ Pacifiquement, cit. 4). *Graves, immenses* (→ Hypertrophier, cit. 2), *redoutables problèmes* (→ Intéresser, cit. 8). — *Problèmes économiques, financiers*. — *Le problème social, les problèmes sociaux* (→ Apporter, cit. 19; enchevêtrer, cit. 5). *Le problème du logement, de l'alcoolisme* (→ Facteur, cit. 3).

L'unique chance de paix, semble-t-il, serait, au contraire, comme le demande l'Angleterre, de ne pas localiser le conflit, d'en faire un problème diplomatique européen, auquel toutes les puissances seraient directement intéressées, et que toutes les chancelleries s'appliqueraient à résoudre (...).

MARTIN DU GARD, les Thibault, t. VI, p. 173.

[a] (Appliqué à une personne). *Avoir des problèmes familiaux, personnels, professionnels*. *Problèmes sexuels*. — *Problèmes psychologiques, ou absolus, problèmes* : conflits affectifs, difficultés à trouver un bon équilibre psychologique. *Adolescent qui a des problèmes, de gros problèmes*. *Il est à l'âge où l'on a des problèmes*.

Avec une valeur générale, couvrant la théorie (→ ci-dessus, 1.) et la pratique. *Étude, traitement des problèmes*. *Science du traitement des problèmes* (problémologie).

♦ 3 Par métonymie. Chose, personne qui pose un, des problèmes (au sens 2); chose incompréhensible. *Pour moi, son attitude est un vrai problème*. → **Enigme**.

Personne qui a des problèmes psychologiques et qui en pose aux autres. *Cet élève, cet adolescent est un problème*.

Annexe 8 (Chap. 5) : articles lexicographiques de évaluer – évaluation.

Article lexicographique extrait du Larousse Lexis :

ÉVALUER [evalye] v. tr. (de l'anc. fr. *value*, valeur, de *valoir*; 1350). *Évaluer quelque chose*, en déterminer plus ou moins approximativement la valeur, l'importance : *On peut évaluer à deux mille francs par jour le chiffre d'affaires moyen de cette boutique* (syn. ESTIMER). *Elle évalue exactement la distance qui la sépare du point où il faudra tourner à gauche* (Mandiargues). ♦ **évaluable** adj. (1845). *Une foule difficilement évaluable*. ♦ **évaluation** n. f. (1361). *Une rapide évaluation du montant des réparations fait prévoir trente mille francs de frais* (syn. ESTIMATION). ♦ **évaluatif, ive** adj. (v. 1960). Qui sert à évaluer : *Recherche évaluative*. ♦ **évaluateur** n. m. (1865). Ce qui sert à déterminer la valeur des choses. ♦ **réévaluer** v. tr. (v. 1950). Procéder à une nouvelle estimation de la valeur. ♦ **réévaluation** n. f. (v. 1950). *Réévaluation de la monnaie. En période d'instabilité monétaire, il est procédé annuellement à la réévaluation des immobilisations et des amortissements*. ♦ **sous-évaluer** v. tr. (1856). Évaluer au-dessous de sa valeur, de son importance : *Sous-évaluer les effectifs de l'ennemi* (syn. SOUS-ESTIMER). *Une monnaie sous-évaluée*. ♦ **sous-évaluation** n. f. (Pl. 2.) ♦ **surévaluer** v. tr. (v. 1900). Évaluer au-dessus de sa valeur : *Les prix des terrains sont surevalués*. ♦ **surévaluation** n. f.

Article lexicographique extrait du Grand Robert :

ÉVALUER [evalue] v. tr. — 1366, *eevaluer*; var. *avaler* (1230) jusqu'au xvii^e; de *es-* (ou *a-*), anc. franç. *valuer*, et suff. verbal.

♦ 1 Porter un jugement sur la valeur¹, le prix de (qqch.). → *Estimer, priser. Faire évaluer un meuble, un tableau, par un expert.* → *Expertiser. Évaluer la fortune de qqn. Évaluer un patrimoine d'après certaines données.* → *Calculer, chiffrer, supputer. Évaluer les diverses parties d'un bien, les divers éléments d'un ensemble.* → *Ventiler. Évaluer un bien au-dessus (→ *Majorer, surestimer*), au-dessous (→ *Sous-estimer, sous-évaluer*) de sa valeur. Sa maison est évaluée sept cent mille francs, à sept cent mille francs (→ *Valoir*). À combien, combien a-t-on évalué son mobilier? On l'évalue approximativement, à peu près à...; on l'évalue à un million au minimum (→ *Au bas mot*), au maximum. Évaluer une marchandise, le cours d'une monnaie.* → *Coter. L'arbitre a évalué le dommage.* → *Arbitrer. Évaluer les dépenses et les recettes à inscrire au budget.* → *Évaluation (du budget).* — Pron. (Possib). *Le déficit peut s'évaluer à cent mille francs.*

Le denier romain, lequel Budée estime du poids d'un drachme, qui est, à raison de huit en (à) l'once, évalué à trois sols six deniers tournois (...)

F. GARRAULT, Recueil des principaux avis du compte par écus (1578).

Nous évaluâmes cette nuit le contenu total du coffre à un million et demi de dollars; et, lorsque plus tard nous disposâmes des bijoux et des pierres (...), nous trouvâmes que nous avions singulièrement sous-évalué le trésor.

BAUDELAIRE, Trad. E. POE, Histoires extraordinaires, «Le scarabée d'or».

Puis, sa pensée déviant malgré lui, il essaya d'évaluer la fortune paternelle et quelle en serait sa part; mais il n'avait là-dessus aucune donnée précise.

MARTIN DU GARD, les Thibault, t. IV, p. 204.

(1870). Déterminer (une quantité) par le calcul (sans recourir à la mesure directe). → *Calculer. Évaluer un volume (→ *Cuber*). Évaluer le débit d'une rivière.* → *Jauger. Possible, impossible à évaluer (→ *Commensurable, incommensurable, incommensurable*).*

♦ 2 Fixer approximativement. → *Apprécier, estimer, juger. Évaluer une distance à vue d'œil, à vue de nez.* → *Juger (au juger), vue.*

Vouzelles et lui jouaient aux échecs sous le bombardement quotidien. Quand les points de chute se rapprochaient, ils interrompaient le jeu, évaluant la distance.

J. CHARDONNE, les Destinées sentimentales, p. 409.

Il ne pouvait souffrir la vue d'une montre et, en fait, il n'y en avait pas une seule dans toute sa maison (...). Il évaluait le temps, et surtout l'heure des repas (...).

CAMUS, la Peste, p. 133.

(Abstrait). Évaluer les qualités, les défauts d'une personne, d'une chose. → *Apprécier, déterminer, juger. Il n'a pas encore évalué les risques, les chances de son entreprise.* — Pron. Il a du mal à s'évaluer.

En fait de sentiments, ce qui peut être évalué n'a pas de valeur.

CHAMFORT, Maximes et pensées, Sur les sentiments, XLV.

Fernando Lucas savait accepter tous les risques dont il avait, depuis longtemps, évalué la quantité et la qualité.

P. MAC ORLAN, la Bandera, XII.

♦ **ÉVALUÉ**, **ÉE** p. p. odj. Ce tableau est évalué à trois cent mille francs. Foule, assistance évaluée à deux mille personnes environ.

DÉR. Évaluable, évaluateur, évaluation. ♦ **COMP.** Sous-évaluer, surevaluer, réévaluer.

ÉVALUATION [evalyasj] n. f. — 1365, *evaluacion*; de *évaluer*.

♦ 1 Action d'évaluer¹. → *Appréciation, détermination, estimation, expertise, prise, comparaison. Évaluation d'une fortune.* → *Calcul. Évaluation des marchandises en magasin.* → *Inventaire. Évaluation d'une chose au-dessus de sa valeur.* → *Majoration (→ *Surestimation, surestimation*). Évaluation de biens mobiliers dans un bilan (→ *Bilan*, cit. 2).* → *Évaluation budgétaire*: détermination des dépenses et recettes à inscrire au budget. — *Évaluation d'une distance, d'une longueur. Évaluation approximative.* → *Approximation. Évaluation d'un débit (cit. 5).* — *Moyen d'évaluation (→ *Échelle*). Procéder à, se livrer à une évaluation.* — *Fisc. Évaluation d'office*: estimation de bénéfices, de profits, effectuée par l'Administration sans procédure contradictoire.

Le peu d'uniformité dans les mesures met continuellement dans la nécessité de faire des évaluations (...).

CONDILLAC, la Langue des calculs, I, 13.

Il ne faut donc pas dire seulement que dans le monde moderne l'échelle des valeurs a été bouleversée. Il faut dire qu'elle a été anéantie, puisque l'appareil de mesure et d'échange et d'évaluation a envahi toute la valeur qu'il devait servir à mesurer, échanger, évaluer.

Ch. PÉGUY, la République..., p. 351.

(...) s'il faut passer une commande aux ateliers de fabrication, il ne sait quelle quantité conserver en stock, évaluation complexe que nul règlement ne prescrit et qui exige le tact du chef.

J. CHARDONNE, les Destinées sentimentales, p. 307.

(...) relevant dans son *Journal* le propos de Byron, selon lequel il n'y a pas plus de quatre mille personnes au monde capables d'entendre les poètes, il (Vigny) est allé jusqu'à former le souhait que sur ces quatre mille il pût y avoir au moins la moitié de lecteurs français. Cette évaluation était assez pessimiste (...).

Émile HENRIOT, les Romantiques, p. 168.

♦ 2 (1691). Valeur, quantité évaluée. *Évaluation insuffisante, trop, faible.* → *Mesure, prix, valeur. Corriger, rectifier une évaluation.*

COMP. Sous-évaluation.

Annexe 9 (Chap. 5) : articles lexicographiques de résultat.

Article lexicographique extrait du Larousse Lexis :

RÉSULTAT [rezylta] n. m. (lat. scolast. *resultatum*; 1589).
 1. Ce qui arrive, se produit à la suite d'une action, d'un événement, de l'application d'un principe, d'une opération mathématique : *Le résultat d'une démarche, d'une négociation* (syn. CONCLUSION, DÉNOUEMENT, EFFET). *Leurs tentatives ont abouti à des résultats inespérés* (syn. RÉUSSITE, SUCCÈS). *Tant de peine et de dépenses n'ont donné que des résultats négatifs, dérisoires, lamentables* (syn. CONSÉQUENCE). *Je refuse d'engloutir des millions pour rien : je veux des résultats* (Beauvoir). *Le résultat d'une expérience de physique, de chimie*. *Mahorin s'était tortillé la barbe en prenant connaissance des résultats de l'analyse* (Bazin). *Le résultat d'une addition* (syn. SOMME), *d'une soustraction* (syn. RESTE), *d'une multiplication* (syn. PRODUIT), *d'une division* (syn. QUOTIENT). —
 2. Réussite ou échec à un concours, à un examen, à une compétition; liste des candidats qui ont réussi : *Le journal a publié les résultats des admissibles à l'agrégation, à Polytechnique*. *Proclamer les résultats d'une composition d'histoire* (= les notes et les places). *Commenter les résultats d'un match*. ♦ n. m. pl. Solde du compte des profits et pertes : *Les résultats de l'exercice ont été bons*. ♦ **résulter** v. tr. ind. (lat. scolast. *resultare*, rebondir, de saltare, saut; 1495) [sujet nom de chose] Être la conséquence, l'effet d'une cause : *Son état de santé résulte d'un excès de travail* (syn. PROVENIR). *L'amélioration du bien-être qui résulte des réformes sociales* (syn. ÊTRE ISSU, PROVENIR); et, impers. : *Que résultera-t-il de toutes ces démarches?* (syn. ADVENIR, ARRIVER). *Il est résulté de grandes pertes de sa mauvaise gestion*. *Telle pression dans les gaz étant produite, certains effets en résultent nécessairement* (Alain). ♦ **résultant**, e adj. (v. 1500). 1. Qui résulte. — 2. *Math.* Relatif au résultat de la composition de tous les éléments d'un système. — 3. *Mus.* *Son résultant*, son correspondant à deux sons réels émis simultanément. ♦ n. m. *Math.* Résultat de l'élimination de la variable entre deux équations. ♦ n. f. (v. 1650). 1. Résultat de l'action conjuguée de plusieurs facteurs : *Cette crise est la résultante des erreurs du gouvernement précédent* (syn. CONSÉQUENCE). — 2. *Math. et Mécan.* Vecteur unique (s'il existe) équivalant à un système de vecteurs glissants. (Des vecteurs concourants ont toujours une résultante : leur somme géométrique appliquée au point de concours de leurs supports.) — 3. *Résultante de transformations (d'opérations)*, transformation (opération) équivalant à l'ensemble de ces transformations (opérations) effectuées successivement.

Article lexicographique extrait du Grand Robert :

RÉSULTAT [rezylta] n. m. — V. 1589 : mot didactique, entré dans l'usage courant au XIX^e : Littré n'en donne que deux exemples, tirés de Voltaire et de Bonnet (*In Pougens*) ; cependant, on le trouve chez Molière (→ *Conférence*, cit. 1 ; et ci-dessous, cit. 1) et fréquemment chez Buffon (→ *Expérience*, cit. 30 ; *imitation*, cit. 4...) : lat. scolast. *resultatum*, dér. de *resultare*. → *Résulter*.

†1 (Langue class.). Ce qui résulte d'une cause. → *Effet*. — REM. Dans cet emploi, le mot a la valeur d'un substantif d'action.

Le résultat de tout est qu'en ami fidèle
Ce soir je vous invite à souper avec elle (...)
MOLIÈRE, *L'École des femmes*, I, 1.

†2 Mod. Ce qui arrive, se produit, commence à exister à la suite et comme effet de qqch. (activité, événement, phénomène, opération, principe). → *Conséquence*, *contrecoup*, *effet*, *événement* (vx), *fin*, *issue*, *solution*, *succès*, *suite*. — REM. *Résultat* se distingue de ces mots en ce qu'il implique un caractère durable (provisoire ou définitif) par rapport aux causes (de même que *conclusion*, *dénouement**, *fin*...) ; en outre, il emporte souvent l'idée d'efficacité, d'utilité (comme *fruit*, *succès*) et s'emploie surtout à propos de l'activité humaine. — Être le résultat de qqch. → *Consécutif* (être), *dépendre*, *ensuivre* (s ; cit. 3), *naître*, *résulter* (de). *Le surmenage, résultat d'un excès de travail* (→ aussi *Accompagnement*). *Le résultat est que...* (→ *Couvrir*, cit. 2). → *Ressortir*, *résulter*, *suivre*... *Quel sera le résultat de tout cela ?* → *Sortir*. *Avoir pour résultat, avoir le résultat de...* (→ *Obscurcir*, cit. 6). → *Entraîner*, *produire*. *Donner un résultat* (→ *Critique*, cit. 23 ; *obstination*, cit. 4). *Conduire**, *mener à tel résultat*. *Le résultat d'une situation, d'un état de choses, des nécessités sociales...* (→ *Juif*, cit. 6). *L'histoire, au lieu d'enregistrer* (cit. 4) *les résultats*...

Comme si les premiers résultats de la Révolution que ces élections commençaient, la suppression de la dime, la suppression de l'octroi et des aides, la vente à bas prix de moitié des terres du royaume, n'avaient pas produit la plus subite amélioration dans le sort du pauvre qu'aucun peuple eût vue jamais !

MICHELET, *Hist. de la Révolution franç.*, I, 1.
Djemylèh (...) s'endormit d'un sommeil profond, résultat naturel de sa grande jeunesse et de l'épuisement de ses forces (...)

J.-A. DE GOBINEAU, *Nouvelles asiatiques*, p. 246.
Ce que produit une activité consciente dirigée vers une fin ; cette fin même. → *Œuvre*, *ouvrage*. *Celui, celle qui a obtenu, produit un résultat*. → *Auteur*, *cause*, *ouvrier* (L., 3) ; *capable*, *efficace*. *Rechercher un résultat, viser à un résultat* (→ *Châtier*, cit. 5). *Atteindre le résultat* (→ *Aboutir*, *réussir*), *obtenir deux résultats* (→ *Faire coup** double). *Ce qui sert à obtenir un résultat*. → *Instrument*, *moyen*, *opération*. *Approcher du résultat*. → *Progrès* (cf. *Toucher au but*). *Il faut des années de patience et d'efforts pour obtenir un résultat* (→ *Échec*, cit. 12), *pour arriver à un résultat* (→ aussi *Alternance*, cit. 2). *Le résultat cherché ne valait pas toute cette peine* (cf. *Le jeu n'en valait pas la chandelle*). *Résultat contraire à la prévision* (cit. 4). *Beaux, heureux résultats*. → *Réussite*, *succès* ; et aussi 2. *bien*, *bonheur*, *chance*. *Résultat inespéré, plus qu'honorable* (cit. 16), *satisfaisant*. *Résultats insuffisants, décevants, dérisoires* (→ Une affaire qui tourne en eau de boudin* ; la montagne* qui accouche d'une souris ; et aussi *fatalité*, cit. 3). *Résultats peu décisifs* (→ *Orientation*, cit. 2), *nuls*. *Le résultat négatif* (cit. 13) *des recherches*. *Résultats désastreux, pernicieux* (→ *Méfait*). — *Le résultat d'une entreprise, d'un plan, d'un travail... Voici le résultat de nos discussions, de nos réflexions. Beau résultat !* (iron.).

Le résultat de ses réflexions était toujours que le suprême bonheur consiste à ne rien faire.

ZOLA, *Thérèse Raquin*, XVIII.

Résultats concrets, réels, tangibles.

(Au plur.). Réalisations concrètes. *Nous exigeons des résultats*. *Avoir* (→ *Fructifier*, *réussir*), *obtenir* (→ *Recueillir*) *des résultats avantageux, fructueux*. — *Comptab.* *Bénéfices ou pertes que laisse l'exploitation d'une entreprise*. *Résultats d'exploitation*.

Et dans tout ça, il n'y a pas que des mots, repris M. Scharbeck. Il nous arrive bel et bien avec des résultats.

J. ROMAINS, *les Hommes de bonne volonté*, t. V, XXII, p. 192.

Sans résultat. → *Infructueux*, *stérile*. *Faire qqch. sans résultat* (→ *Cuisiner*, cit. 4). — (Par oppos. à *intention*, *projet*). *Il n'y a que le résultat qui compte, le résultat seul compte* (cit. 40). → aussi *Calculer*, cit. 6. — (En parlant d'une opération mentale). *Résultat(s) d'un raisonnement, d'une discussion, d'un entretien*. — Sc. *Le résultat d'une activité scientifique* : les éléments positifs qui en résultent. *Résultat d'une expérience* (cit. 16). → *Exprimer*, cit. 46 ; *procès-verbal*, cit. 4. *Résultat de recherches*. *Les résultats de la science* (→ *Philosophie*, cit. 8).

(En parlant d'une chose concrète produite par une opération). *Le résultat d'un mélange* (cit. 1). *Le déjeuner périgourdin* (cit.), *résultat de méditations savantes*. Loc. adv. (Rare). *En résultat* : en fin de compte.

(Ellipt., en tête de proposition). Fam. *Il a voulu sauter par la fenêtre ; résultat, il s'est foulé la cheville*. → *Total* (fam.).

†3 *Solution* (d'un problème). *Il connaissait le résultat d'avance*. — Phase ultime d'un calcul ; le troisième élément (→ *Composé*) associé à un couple dans une application, une opération arithmétique. → *Produit*, *quotient*, *reste*, *somme* ; et aussi *réponse*. *Résultat d'une opération*.

†4 (xx^e, dans une compétition). *Admission ou non-admission, réussite ou échec* (à un examen, un concours) ; liste de ceux qui ont réussi. *Affichage, lecture, proclamation des résultats*. *Il a passé l'oral hier, mais il n'a pas le résultat*.

Il réussit assez bien cette fois, à l'écrit. Mais ce fut dur d'attendre les résultats de l'admissibilité (...) Le jour du résultat final arriva. On affichait dans la cour de la Sorbonne les noms des candidats reçus.

R. ROLLAND, *Jean-Christophe*, Antoinette, p. 901-902.

« Asseyez-vous, messieurs », disait le préfet des études qui prenait un air solennel. Et alors le professeur lisait les résultats de la dernière composition. Quel instant ! Premier : Lénier (Joanny).

Valéry LARBAUD, *Fermina Marquez*, VIII. (1885, in D. D. L. : *résultats partiels*). *Les résultats connus de l'élection* (→ *Gagner*, cit. 52). *Résultats partiels, provisoires, définitifs*. *Entendre à la radio, lire dans les journaux les résultats du scrutin*. — Sports. *Résultats d'un match* (cit. 2), *d'une compétition*. *Les résultats des courses*. Loc. fam. *Le résultat des courses* : le résultat. *Est-ce que tu es reçu ? Je n'ai pas encore vu le résultat des courses*. Alors, cette réunion ? Quel est le résultat des courses ?

La dimension argumentative des collocations textuelles en corpus électronique spécialisé au domaine du TAL(N)

Estelle DUBREIL

Résumé

Traditionnellement, l'association lexicale préexistant au phénomène collocationnel est qualifiée d'arbitraire. Or en soumettant les collocations textuelles – CT – à l'analyse sémantico-discursive des possibles argumentatifs – PA –, modèle de description de la signification lexicale, non seulement cette association s'avère bien souvent argumentativement motivée, mais parfois prédictible. Aussi, nous postulons qu'une CT correspond potentiellement soit à une collocation argumentative, soit à une relation argumentative, soit enfin à une collocation arbitraire. Cette triple hypothèse a été testée sur le corpus électronique spécialisé TAL(N), grâce à une méthodologie pluridisciplinaire adaptée au caractère électronique des données (outils logiciels et Internet). Les résultats obtenus valident notre hypothèse par la présence équilibrée de collocations argumentatives et de relations argumentatives, comparée aux collocations arbitraires. Ces résultats nous confortent dans l'idée de reconsidérer le caractère arbitraire du principe d'association collocationnelle et nous permet d'en proposer une description en termes de blocs d'argumentations externes.

Mots-clés : collocation textuelle, possible argumentatif, motivation argumentative, collocation arbitraire, signification lexicale, bloc d'argumentation, corpus électronique, communauté TAL(N), linguistique de corpus

Abstract

Traditionally, the lexical association that leads to the collocational phenomenon is thought to be arbitrary. Nevertheless, when applying the semantico-discursive theory of the Possible Argumentative – PA – on Textual Collocations – TC – this association arises as argumentatively motivated and possibly predictable. We state that TC possibly corresponds to an argumentative collocation, an argumentative relation or an arbitrary collocation. This triple assumption has been tested on the specialized electronic corpus TAL(N), with a multidisciplinary methodology adapted to electronic data (Software and Internet Tools). Our hypothesis has been validated by the results we have obtained. Indeed, they highlight the equal distributions of argumentative collocations and argumentative relations within the corpus compared to arbitrary collocations. This reinforces our intuition of reconsidering the arbitrary nature of collocational associations and motivates our proposition for a description model for composing collocational associations in terms of external argumentative blocks.

Keywords: Textual Collocation, Possible Argumentative, Argumentative Motivation, Arbitrary Collocation, Lexical Signification, Argumentative Block, Electronic Corpus, TAL(N) Community, Corpus linguistics